UNCLASSIFIED

AD 243 000

*Reproduced by the*

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

20030715073

UNCLASSIFIED

# CONTROLLING LITERATURE BY AUTOMATION

277

*To Be Presented at the*



## IV ANNUAL MILITARY LIBRARIANS' WORKSHOP

Sponsored by

**ARMED SERVICES TECHNICAL INFORMATION AGENCY**

- *The Scientific Information Center of the Department of Defense*
- *An Activity of the Air Force Air Research & Development Command*

**5-7 October 1960 . Washington D.C.**

## FOREWORD

The first seven chapters in this compilation were prepared by the authors as term papers in a course recently given by American University on "Problems in ADPS." The authors, as students in the course, represent a cross-section of ASTIA leaders; hence, their papers represent a substantial coverage of the plans, the problems, and the installation of automation in the Armed Services Technical Information Agency (ASTIA). As such, it was decided that use of the papers as workshop material was ideal.

The papers by Mr. Calvin Mooers and Mrs. Claire Schultz are added because they are so closely related to ASTIA's automation as to be considered appropriate. They have been added with the consent of the authors who will discuss them at the Workshop. Both authors have been associated with the development of ASTIA's automation.

It is hoped that these records of the actual experiences of ASTIA in automating its major functions will be helpful to others as progress is made toward improved utilization of the ever increasing masses of scientific and technical information.

WOODROW W. DUNLOP
Colonel, USAF
Commander and Director

TABLE OF CONTENTS

Page

SEQUENCE OF TERM PAPERS

## Chapter 1

### Preparing for Automation
by Lt Colonel William Hammond, USAF, ASTIA

We, of ASTIA, who are attending this course played key roles in developing an automation system. As managers we also had to recognise and solve the management problems associated with converting our plans into the actions required to install the system. We have prepared a coordinated series of presentations dealing with some of these management problems. We will touch only lightly on the details of our operation and of our computer applications. For those of you who are interested in such details, we suggest you obtain a copy of AD 227,000, Automation of ASTIA, A Preliminary Report, dated 1 December 1959.

We are not presenting our approach as the blueprint for automation. We hope only to contribute to the reservoir of knowledge and experience being accumulated as each new automated application breaks through the manual barrier. To quote Mark Twain, "We should be careful to get out of an experience only the wisdom that is in it----and stop there".

My subject is, "Preparing for Automation" - the gearing-up period. It covers formulation of our objectives, how we went about setting up the implementing machinery, and how our implementing actions were conceived and executed.

The Armed Services Technical Information Agency --ASTIA for short-- is a unique organisation. It is assigned to the Air Research and Development Command, but it has a tri-service responsibility. Its policy stems directly from the Department of Defense. ASTIA is specifically charged with the timely exchange of scientific and technical information within the Department of Defense research and development community.

ASTIA has under bibliographic control about three quarters of a million technical reports accumulated since World War II. Only 200,000 of these reports are included in the current automated operation. They are the 200,000 most recently cataloged into the collection and represent the documentation of the past seven years' military research and development program. About 97% of ASTIA's services are currently being provided from this collection.* More than a hundred new reports a day are added to the collection.

ASTIA services about 2,000 military and government organizations, and over 3,000 contracts, spread among more than 1,300 civilian contractors. These services consist of processing over 25,000 new reports a year into the system and announcing their acquisition in a twice monthly Technical Abstract Bulletin; furnishing one-half million copies of technical reports a year to fill requests, and providing 1,600 report bibliographies covering any subject or combination of subjects in the collection.

About half the documents requested, some 250,000 a year, are not carried in stock and must be reproduced from microfilm.

ASTIA's approach to automation was pretty much along the lines suggested by Canning and Chapin—except for timing. Early considerations covered several years and dealt mostly with cost factors and technical feasibility—and waiting for the state of the art to advance. Experts in all phases of the operation were called upon. Equipment manufacturers prepared studies and submitted proposals; ASTIA management prepared still more studies and evaluated the accumulation of data.

*The ASTIA Document Collection (AD reports)

The decision to automate was reached in mid-November 1958. Within a week the mass of accumulated data was ground into the format required by AF Manual 171-9, Management of Data Processing Equipment, for submission of a proposal to automate. In an attempt to expedite higher Command action, the equipment vouchers were submitted along with the proposal. This was a mistake - it actually confused the issue, and delayed coordinating actions by higher headquarters staff.

The major consulting service involving the computer applications had been provided by the IBM Corporation—and as you might expect, IBM equipment had been recommended. However, to ASTIA's credit, consideration of other manufacturers equipment continued.

The actual gearing up for automation began in earnest the day the decision to automate was made, some seven months before higher Command approval. The ASTIA Commander's first action was to name a project officer and designate the major organizational segment that would assume primary staff responsibility. The project officer was charged with developing the detailed implementing plans and the organizational machinery for executing the plans.

I was selected as the project officer. Prior to my selection, I had been given about four months to study the ASTIA operation and to become familiar with the manual operational procedures. My only association with ASTIA's automation program during that period was to put together the original proposal to automate. This was primarily an editorial and format review; no creative input was involved. My primary guidance came from the AFM 171-9.

About three weeks after the automation proposal was submitted, an EDP Planning Group was established. This group was composed of the heads of the major operational and support elements of ASTIA. The project officer became the chairman of the group. The Research Requirements Officer served as technical advisor. To maintain continuity and to broaden the base of participation, the number two man of each of the major organizational segments served as an alternate member.

The planning group as a whole met once a week. This regular weekly meeting was primarily to monitor the scheduled actions and to keep members abreast of progress. Minutes of these meetings formed the official records of progress toward implementation.

The planning group first undertook the task of establishing a firm planning base. A complete review of the proposal was made and the specific objectives identified in terms of end products, functions, and effective dates.

Our broad general objective under automation was to provide more timely service. In more specific terms:

1. To fill requests for documents from stock in three working days.
2. To fill requests for documents not carried in stock in five working days.
3. To improve inventory management so that 80 to 90 percent of documents requested would be prestocked, and thus filled in the minimum of three days.
4. Provide bibliographic service (by list of ASTIA Documents ) in three working days. (This was later changed to complete printout of all bibliographic and abstract information for each document included in a bibliography.
5. Maintain accurate document accountability.
6. Announce newly acquired documents in the ASTIA Technical Abstract Bulletin within 30 working days.
7. Provide quarterly and annual cumulative indexes to the Technical Abstract Abstract.

To accomplish these objectives it was determined that the following automated applications were necessary:

1. Automatic request validation to determine the need-to-know and security clearance of the non-military requester.
2. Automatic inventory control.
3. Automatic document accountability.
4. Automatic identification of documents requested without reference to specific ASTIA catalog numbers.
5. Automatic duplication check of newly acquired documents.
6. Automatic index preparation for the Technical Abstract Bulletin.
7. Automatic search and retrieval of information held in the ASTIA document collection.

Most of the applications readily identified were associated with the processing the requests for documents. These were "business" or general purpose applications. The applications associated with information retrieval were not so readily identified. We tried everything from brainstorming to expert consultation. We also relied heavily on ASTIA's own research and experience in this area over a period of several years. At this stage in our progress we had to settle for establishing only the short range objectives for the information retrieval application.

After we had identified the automated applications, we next concentrated on the inputs that would be required. We identified the administrative and bibliographic files involved, and the preliminary processing that would be required prior to conversion of these files into machinable form. We determined those areas where files did not exist or where they could not be used in their present form. The creation of an authoritative Thesaurus of descriptors to be employed in the automated retrieval system is an example of one of the major tasks of this nature that was identified.

It was next necessary to determine the resources required to accomplish our objectives. The Commander had established the concept that there must be no reduction in ASTIA's services and that no backlogs could accumulate during the conversion from the manual to automated operation. This required an evaluation of the magnitude of each task to be accomplished and ASTIA's ability to accomplish within the resources to be made available during the period of implementation.

The next phase in developing the planning base was to refine the flow charts of the automated operation and establish a master schedule for implementation. (See Attachment ) Before these actions had been accomplished, it became apparent that the equipment* originally recommended would not provide us, within the resources that could be made available, the end products we desired in accordance with the timing we had established.

By this time we had become very much interested in the Remington Rand new UNIVAC Solid State computer (USS90) which had been announced subsequent to the submission of our proposal to automate. Only the punch card configuration of this computer that had been announced, and a punch card operation would not do the complete job we had in mind. However, further examination of the USS90 computer convinced us that with the addition of magnetic tape and random access capabilities, it would do the job we had set out to do —both the general and special purpose applications that we had envisioned. Additionally, a closer examination of the potential of this computer led us to re-evaluate both our short range and long range objectives in mechanised search and retrieval of information. You will hear more on this subject from Mr. Barden in reference to our plans for integrated data processing.

*IBM 650 RAMAC with magnetic tape for "Business" application and the IBM 9900, Special Index Analyser for information retrieval.

While the Office of the Secretary of Defense was still evaluating our proposal to automate--perhaps with some skepticism--in view of our proposal being based on providing better and faster service at somewhat higher cost—we submitted new equipment vouchers for the Remington Rand USS90 computer. We supported our change in the selection of hardware on the basis that we could move faster toward both our short and long range objectives, and at less cost than with the equipment that we had originally recommended. This was also true for any other hardware then on the open market. On this basis we sold a bill of goods—On 24 June 1959, Headquarters USAF gave its final approval, with an on-the-air date for the USS90 card configuration of 15 February 1960; the magnetic tape configuration 1 July 1960; the RANDEX* to be added on or about 1 October 1960.

We did not receive the official notification of the approval until 7 July 1959. However, on the basis of the advance knowledge of the USAF approval, we were able to process a contract on 29 June for the file conversion required for the business application of our operation. Thus we were able to use funds we had withheld from lower priority contractual projects during the year. No additional funds have to this day been provided by higher headquarters to implement the automated operation. Only the equipment rental has been assumed by ARDC. If we had not been able to process this contract, it is unlikely that we could have automated during this Fiscal Year. Our detailed advanced planning had paid big dividends from the very beginning.

With the official approval in hand in early July 1959, implementation tempo increased. The Data Processing Branch was established and programming and file conversion got under way. The EDPS Planning Group continued to function as the

*Remington Rand trade name - Random Access Unit staff monitoring medium. Project officers were assigned to direct each of the major implementing actions that the group had identified.

The Univac Division of Remington Rand furnished one systems analyst and one programmer full time to assist in developing the computer programs. This service was included in the file conversion contract. In addition to the services covered in the contract, Remington Rand Univac provided technical assistance and the facilities of its Service Bureau on an on-call basis. Although the equipment contract provided for 40 hours of debugging time, all the time that we could profitably utilize was made available as we needed it.

During the implementing period, the Deputy Director of ASTIA usually attended the weekly meetings of the EDPS Planning Group, and participated in the actions of the group. When Command approval was required for design or format of end products, or for any deviations from the approved plan of action, the project officer would schedule a meeting with the Commander to present the factors bearing on the matter under consideration. A sort of informal steering group consisting of the project officer, the research requirements officer, the chief of the Management Division, and usually the chief of the operating division most concerned participated in these meetings with the Commander.

The Commander reviewed the minutes of the weekly meetings of the EDPS Planning Group, and was usually aware of the actions that were on the way up. There was no instance of major disagreement on significant issues among the elements of the ASTIA staff. Command decisions were forthcoming on a timely basis. The Commander maintained a deep personal interest in the automation program and insisted that implementing actions be completed on schedule.

Many ASTIA people made significant contributions to the successful conversion from the manual to the first stage automated operation. This conversion was made on the date scheduled, without the benefit of parallel manual operation. This paper was prepared just four weeks after the operation was put on the air. Although the data cleanup has been challenging, and the workloads the highest in the history of ASTIA, the first stage operation is approaching routine. We are now reprogramming for tapes and developing new programs for the added applications of the second and third stages of automation scheduled for 1 July and 1 October 1960 respectively. The pressure increases daily, but all of ASTIA is behind the effort. The Data Processing activity is operating in a permissive atmosphere.

Chapter 2

## Selection, Training and Relocation of Personnel

by Leroy R. Barnes

### General Situation

1. Colonel Hammond has briefly outlined the get-ready phase of ASTIA's automation program. The subject of this paper is personnel--selection, training and relocation. As in any conversion from essentially a manual to an automated operation, the personnel aspects of the conversion are both widespread and, in many instances, a very personal, close to home matter. I doubt if there is a single employee in our Agency who feels he or she has been unaffected by our automation program. As a matter of fact, there are few if any persons who have not been affected even if only slightly. For example: The mail clerks now process requests in punched card form; now that we have inventory control, the document stock clerks only handle those requests for which there are documents on the shelf; the supply clerk is now ordering, stocking and controlling many new items of supply and equipment.

2. I could go on with many more examples of the effect of automation upon people in all organizational units. Suffice it to say that everyone is affected, duties are changing, and the old day-to-day routines are no longer applicable. We have been told all of our lives that people resist change--that they do not want their daily routines upset. This is generally true - and is particularly true if they are not informed of the nature of the change and cannot personnally estimate the impact. In ASTIA we made a very definite point of keeping our personnel informed. It is the Commander's policy to inform the people of planned actions and keep them informed to the greatest extent possible.

### Keeping Employees Informed

3. On 15 September 1958 a memorandum, subject: "IBM Training", was provided All Supervisors and Employees of ASTIA from the Commander. This date

was 17 months ahead of the automation operational date and 20 months or more ahead of any possible adverse employee action. In this memorandum employees were urged to take an aptitude test, and to take advantage of local training courses to fit them for the new positions in the automated operation. The point I am making is this--ASTIA employees were informed of the automation plans at the earliest possible date; they were kept informed; and most important, they <u>were</u> given the opportunity to participate and <u>were</u> selected for these new positions. The results have been dramatic--we have not only gone forward in a "permissive" atmosphere as pointed out by Colonel Hammond, but it also has been and is an enthusiastic atmosphere. To counter any raised eyebrows, I offer these facts: (1) During the past year and a half, there has been no appreciable change in the total number of ASTIA employees--<u>346</u> in September 1958 and <u>352</u> today. For one short period we did have approximately 360 personnel assigned. (2) Large backlogs both in document processing (cataloging, abstracting, announcement) and in request processing were eliminated. (3) The major personnel effort (25 man-years) in the preparation for automation was accomplished in house--the development of the ASTIA Thesaurus of Descriptors and the assignment of the descriptors to the AD collection (the subject of Mr. Rehbock's paper); the training of personnel; the development, testing, and debugging of the computer programs; and the many other time-consuming aspects of getting ready for automation.

4. I submit that this all-out effort reflects enthusiastic support and an amazingly high degree of personal and unselfish effort on the part of employees at all levels.

5. I believe that the success of this program is directly attributable to the participation of our experienced personnel, and it is my firm conviction that we could not have hired enough EDPS specialists as such to have progressed this far.

Phase Out

6. To provide a partial insight into specific aspects of the staffing for EDPS, it is pertinent to briefly describe the kinds of employees in ASTIA. In the Document Processing Division, Scientific Analysis Branch, the personnel perform a subject analysis of the reports and review an author abstract or prepare an abstract. In the main these people have scientific or technical backgrounds--Mathematics, Physics, Engineering, etc. The Civil Service people have had difficulty associating these qualification requirements with the job title category of Librarian. In the Customer Service Division some of the employees have a background similar to those in the Document Processing Division with somewhat more emphasis on the Reference Librarian profession. Since these people are directly associated with our customers in their day to day work, there is the requirement for a high degree of public relations competency. The Reproduction Division is staffed with personnel experienced in printing, photographic and related fields. Throughout the entire organization there is a fairly high percentage of personnel with varied backgrounds- no specific pattern - who by long service in ASTIA have become qualified "Documentalists", a designation that is gradually becoming fully recognized.

7. A "phase out" study was completed in October 1959. This study identified, in specific detail, positions associated with those functions to be automated. This survey did not deal with related administrative and supervisory positions that would be affected. There were 25 operational positions affected.

Staffing the Data Processing Branch

8. With few exceptions, the personnel in the Data Processing Branch were reassigned from within ASTIA. They brought to their new jobs a wealth of

substantive program knowledge and experience. The concentration of experience, the timeliness of training, the thorough systems analysis, and pre-operational program development, testing and debugging have largely compensated for the lack of numbers.

9. In the initial proposal, the manning requirement for the Data Processing activity was 27 spaces and an estimated eventual savings of 70 spaces based on estimated requirements in a manual operation. With the change from IBM to Rem Rand equipment predicated upon certain conceptual changes, the manning requirement changed to 17 civilians plus 1 military for the initial operation - Phase I (punched card operation) - with an additional 3 civilians under Phase II (magnetic tape) beginning 1 July 1960. Subsequently with added workloads and program changes, a requirement of 3 additional civilian spaces was recognized for a total of 23 civilians and 1 military. A manning chart is included as TAB A.

10. To provide a basis for screening, selection and training, three aptitude tests were given ASTIA employees with the following results:

| Test | Number Participating | Range of Scores |
|---|---|---|
| Aptitude test for EDPM Programmers | 18 | 67-A - 4D |
| Punch Card Machine Operator Aptitude Test - IBM | 65 | 41-A - 1D |
| Card Punch Aptitude Test (The Psychological Corporation) | 65 | 164-A - 57D |

Although the test results were not conclusive in all respects, there were indicators which proved very helpful in making selections.

11. The 3 assigned programmers scored 67 A, 59-B and 53-B, 3 of the 4 best scores in the programmer aptitude test. The background, training and experience of our programming staff are significant. The senior programmer has a master's degree in mathematics, 4 years of bibliographic experience in ASTIA and six months previous experience as a programmer trainee. Another

programmer has an accounting background with several years of administrative experience in ASTIA. Our third programmer is a Business Administration major who worked part time in ASTIA while obtaining his degree from American University and then had better than a year's experience in the management engineering staff prior to his assignment to data processing. The console operator had several years' experience in the report processing (cataloging) function in addition to communication equipment coding and operating experience. The programmers and console operators double in brass in performing the programming and computer operations. This arrangement has proven invaluable in the development of their proficiency.

12. Another factor that contributed to the success of our staffing was the arrangement whereby we selected the best of the card punch operators from the Rem Rand employees working on the service contract. The timing was ideal--we employed them just as the contract was being wound up!

13. A brief summary of the relocation or reassignment of ASTIA employees to the Data Processing Branch is pertinent. The Branch Secretary, the PCAM Supervisor and a Programmer were transferred from the Plans and Analysis Branch, Management Division; two Programmers were transferred from the Customer Service Division; the Console Operator was transferred from the Document Processing Division; and the PCAM (Key Punch, Tabulating, Clerk-Verifiers) Operators were transferred from all three operating divisions.

14. I have approached this task of describing ASTIA's experience in selecting and training a data processing staff with tongue in cheek. I believe, without qualification, that the success of the program is mostly the result of exceptional staffing.

15. ASTIA's "data processing manager" has over 20 years of military service with the Army Engineers and the Air Force primarily in geodetic and cartographic work. He has command and staff experience in the Army, Navy and

Air Force. In the data processing field he was an experienced user of special purpose data processing equipment associated with his professional field on such projects as establishing the instrumentation grid for the Cape Canaveral missile range and for the Special Weapons Command test range in Nevada. His last job before coming with ASTIA was to establish the geodetic datum for missile operation from Formosa against targets on the mainland. These jobs are significant only in that they required task force type organizations and Airborne electronic survey equipment. Although he had no previous experience in technical documentation or with ASTIA's operation, he had about 5 months to study the ASTIA operation prior to his assignment as project officer. It took most of that time to become familiar with the manual operational procedures and to become conversant in the technical and procedural lingo.

16. The Deputy Chief of the Data Processing Branch has 25 years of first hand line and staff experience in date processing. He has a combination of government and industrial experience. He is the only principal member of the organization recruited from outside the Agency. Prior to joining our staff he was on the Hq USAF staff and was largely instrumental in steering ASTIA's proposal through USAF and DOD. He brought to our operation a wealth of experience, know-how, and savvy and has made an immeasurable contribution to the success of the program.

17. I could name many others in ASTIA who have made great contributions to the success of the project--this has truly been a team effort. An example is the PCAM Supervisor who is a Princeton graduate with a Foreign Affairs major and three years experience as a translator in addition to eight years in a variety of assignments in ASTIA. We discovered very early in our operation that the requirements of this position were much beyond being just a "supervisor" of machine operations. There was the requirement for a person with detailed knowledge and experience in request processing - the function being automated.

Training

18. The training of ASTIA Employees for EDPS falls into two general categories: (1) specific training for programmers, console operators, PCAM operators and (2) courses offered locally by universities such as this one - "Problems in ADPS".

19. In the first category, our programmers and console operator have all had a concentrated four weeks course in basic programming and a two week course in Flowmatic Programming. Three other employees - one from the Reproduction Division, one from the Customer Service Division and one from the Plans and Analysis Branch, Management Division - took the Flowmatic Course. The PCAM operators reassigned from within the Agency were all given specialized training by Rem Rand in the machine operations.

20. In the second category, selected personnel from all parts of the organization are taking after duty courses to learn more about EDPS. This fiscal year a total of 32 ASTIA employees will have completed the course "Automatic Data Processing Systems" and 7 are taking this course, "Problems in ADPS" provided here at Arlington Hall by American University.

21. Through the years ASTIA has had a limited but quite significant career development and training program. For example--two of our key people have had the AMA 4 week Management Course, and we have provided specialized courses to large numbers of our personnel in "Documentation" and "Management of Scientific Information". This year our program has been almost entirely devoted to EDPS training. Incidentally, our training program this year is being funded at about $3,000.00.

Summary

22. At this point in time, the situation is estimated as good or better. ASTIA's automation is on schedule and it was an ambitious schedule. We were on the air in less than 2 years from the beginning of the feasibility study.

In the month and a half of operations, we have overcome every obstacle - there has been no interruption of operation or service to our customers. The flow of work is beginning to even itself out and service time is almost back to the immediate pre-automated operation time which was the best in ASTIA's history. This has been accomplished in the face of ever increasing workloads - in March we received over 50,000 requests for documents - the highest number in the history of the Agency.

23. The question of the impact upon our personnel has been raised. The initial impact is past - by far the majority of our personnel have made the adjustment. Many of them are in new positions with greatly expanded earning potential. There remain a few (about 10) employees who are occupying positions that are in the process of being phased out. They have not as yet qualified for new positions that are developing. We have not had nor do we anticipate the need for a reduction-in-force program because of conversion to automation.

24. The future looks bright from the personnel viewpoint. Within the approved grade structure, there are promotional opportunities for the next several years in the Data Processing Branch. Within the very near future, we plan to implement an approved CSC training and promotion program that will permit hiring a limited number of carefully selected trainees for future staffing in the Data Processing Branch as well as other areas in the Agency.

25. There is no simple formula for the successful handling of personnel aspects of conversion to automation. It is a combination of many things. Listed in order of importance:

(1) A manager with both command and staff experience with the ability to get a job done on schedule

(2) Supporting staff experience

(3) A planning group made up of executives and managers high enough in the organization to reach conclusions and assign resources to get a job done

(4) Selection of the key personnel for the Data Processing Branch from the staff and operating elements

(5) Timing - training, program development, testing, debugging.

26. In summary - it can be stated that by informing your personnel at the earliest possible date, by making possible their participation, by selecting the people from within the organization, by providing for their training for their new assignments - these are the elements that spell success in a conversion to automation.

MANNING CHART

Data Processing Branch
Management Division
Armed Services Technical Information Agency

6 April 1960

| | | Positions | Authorized Manning | |
|---|---|---|---|---|
| TDMD | Branch Office | | | |
| | | Chief | 1 | |
| | | Deputy Chief (Supv. Dig. Comp. Systems Officer) | 1 | |
| | | Digital Computer Systems Analyst * | 1 | |
| | | Secretary (Steno) | 1 | |
| | | | | 4 |
| TDMDD | Data Preparation Section | | | |
| | | Supervisor *Additional Duty | | |
| | | Key Punch Supervisor | 1 | |
| | | Key Punch Operator | 4 | |
| | | Clerk-ICAM Verifier | 3 | |
| | | Tabulator Operations Supervisor | 1 | |
| | | Tabulator Operator | 3 | |
| | | File Clerk | 2 | |
| | | | | 14 |
| TDMDP | Program Section | | | |
| | | Digital Computer Programmer | 3 | |
| | | Clerk-Stenographer | 1 | |
| | | | | 4 |
| TDMDC | Computer Section | | | |
| | | Digital Computer Systems Operator | 2 | |
| | | | | 2 |
| | Total Manning | | | 24 |

TAB A

## Chapter 3

### File Conversion, Data Cleanup and Inventory Control
by Fred A. Keller

#### FILE CONVERSION

In order to understand the extent of the conversion tasks, one must be familiar with the size of the ASTIA collection and the variety of files that had been established over a period of years. During this period of approximately 15 years, ASTIA's collection had grown to over 750,000 individual reports. Currently, the accessions amount to approximately 30,000 new reports annually.

The files that are needed to permit proper servicing and control of such a collection are as follows:

a. Source - record of each report filed by source, fully identifying each report.

b. Validation File - record of subject categories, current security classification and release limitations arranged by AD (ASTIA Document) number.

c. Security Control File - record of current security classification and authority therefore of each report in the collection.

d. Catalog - record of each report by subject, personal author, project or contract number as applicable.

e. Inventory - record of number of copies of classified reports on hand.

f. Field of Interest Register - record of each non-military user's eligibility for ASTIA services - includes subject areas and classification authorized, contract number and date of expiration of the contract.

The ASTIA collection consists of several different series. In preparing for automatic data processing, it was decided that only the most recent series, i.e., AD, warranted the effort involved in file conversion. Even on this basis the task was of staggering proportions. For example, the source file averages one and one half cards per report, and the AD series consisted of about 200,000 reports at the beginning of 1960. The source file requires up to 89 alphanumeric characters of information for complete identification.

The validation file and the inventory file for the 200,000 reports were combined and required up to 30 numeric characters of information each, of which about one-third were non-standard punching as explained in the discussion of the field-of-interest register file. The inventory information had to be added to the validation file on a carefully controlled basis just prior to the operational date for EDPS. The task of taking the inventory is described later under Inventory Control. The security control file, being an authority file which continually required manual access, was not converted.

The catalog consisted of about 1,500,000 catalog cards filed by source, subject, personal author, project number and contract number. This file had been manually maintained and was growing at about 200,000 cards per year. This file had to be replaced by approximately 1,800,000 punch cards representing descriptor-document combinations for use in the machine retrieval system. These punched cards required an average of about 20 numeric characters each.

The Field-of-Interest Register (FOIR) file involved only about 4,000 individual contracts for some 1,300 contractors. However, it was one of the most difficult to convert because a user's need-to-know could encompass any one or more of some 240 subject categories. In order to effectively reflect the subject categories in one punch card, it was decided to use dedicated positions in about 70% of the card. This entire area therefore required non-standard punching. A seven digit code was established for each user. The first four digits represent a specific user and location, and the last three digits represent a specific contract. In addition, a "via" office must be designated for each non-military user who is authorized access to classified information. The "via" office is represented by a four digit number in the same series as the user code discussed earlier. The FOIR card can require up to 260 numeric characters of information with the average somewhere around 100. As an indicator of the impact of non-standard key punching, the rate for the FOIR cards was 25 man-hours per thousand compared with about 11 man-hours per thousand for the source file which is a more conventional key punching job. This type of FOIR card carried all the information required for processing requests as far as user eligibility is concerned. However, in order to facilitate manual access it was necessary to establish a user code-user address file (averaging about 5 cards per user and about 50 alphanumeric characters per card) arranged according to user code and another arranged alphabetically by user name. Finally, it was found necessary to establish a user code-contract number file (averaging about 20 numeric characters per card) arranged by contract number and according to user code in order to facilitate manual access to this information.

The conversion of all these files was accomplished under contract and timed to permit ASTIA to become operational with EDPS on the date initially scheduled.

## DATA CLEANUP

The scope and extent of the data cleanup were recognized early in the planning for automation. Insofar as possible, data cleanup was accomplished in advance of file conversion. In some instances it had to go along concurrently with conversion.

ASTIA's task in the area of data cleanup has been a difficult one because of the variety and extent of its working files plus the fact that, in automation, some files would be combined which had previously been separate and vice versa. The source file was a tremendous task because of an almost unbelievable diversity in report numbering systems that were in use. Contract numbers (which are incorporated into the source file) also represent many different concepts in terms of what is significant in a contract number. In order to assure accuracy and consistency in the punch card file, it was necessary to establish ground rules for standard types of entry in these instances. Then a source file had to be meticulously checked for accuracy of information; report numbers and contract numbers had to be edited reflecting exactly how they were to be key punched in accordance with the ground rules.

Some items of information were common to more than one file. As the key punching progressed, machine runs were made for correlation of the data. Wherever there were discrepancies (and there were many), it was necessary to go back to the source of the information in order to ascertain the correct information.

Data cleanup in the validation file required standardization of distribution limitations. These had been in use for many years. During this time the philosophy of distribution limitations had undergone a number of changes with the result that in the manual file there were many cases in which a variety of limitation notices meant the same thing. Consequently, after standardizing the limitations, it was necessary to screen the entire file and annotate the correct distribution code for each report prior to key punching.

In another aspect of data cleanup, the data processing system was brought into action. It was known that there were numbers in the AD series for which the actual reports had been cancelled or superseded. In other instances the numbers had never been used. Following completion of the validation file, a machine run was made to print out all numbers for which there was no validation information. This list was then checked against the original source file and the microfilm file, and it was found that in some cases reports had been entered into the system without being fully cataloged. In others, there had been cancellations without reflecting the action in the shelf list (numerical file). In view of ASTIA's split operation from March 1953 through January 1958, it is not surprising that such a situation had developed. From a management point of view, the notable aspect is that in the manual operations there had been no effective way of systematically bringing these discrepancies to someone's attention for correction. Occasionally an isolated instance would be discovered by accident and corrected. But that was as far as it went.

Again the FOIR file proved to be one of the most difficult. Some users had been established under a variety of codes and addresses covering the identical location. This situation had to be rectified, and it took one employee six weeks to develop the needed standardization and reflect the changes in the manual file, the punch card file and the addressograph plates.

Finally, it was found that codes for "via" offices were in a different series from those for users. Since many of the "via" offices are also users, it was decided to assign codes in the user series. This action required many man-days to accomplish, but it will pay dividends from now on.

Naturally, a certain amount of data cleanup is still going on and can be expected to continue. The important point here is that the data processing system, itself, can be brought to bear to keep it clean.

## INVENTORY CONTROL

At present the most economical means by which to achieve vastly improved service in processing requests is to increase the ratio of requests filled from stock versus the requests that must await reproduction. Historically about 45% of requested reports have been available from stock and about 55% have to be reproduced. The time depends entirely on the reproduction workload and the types of reproduction that can be used. The time required for reproduction may not be appreciably reduced, but the percentage of reports available from stock can be vastly improved. To achieve this improvement, it is most important to develop a mechanized inventory control in conjunction with the EDPS.

Stock levels and activity (usage) statistics must be maintained for each document in order to provide for proper inventory management. These statistics and stock levels must be immediately available preferably on a random access basis in order to determine the availability of a requested document, to reduce lead time by taking maximum advantage of quantities which may already be in reproduction, to eliminate a shelf check for non-existent stock and to determine optimum re-order quantity if reproduction is required.

It should be kept in mind that under the manual procedures all user requests except for those documents that were known to have been destroyed because of age were sent direct to storage areas, and a shelf search was conducted. Requests for documents which were not available from stock were then forwarded for reproduction to service the particular request or additional requests for the same report in the event they were discovered during the film stripping process. In the summer of 1958 an inventory control was established to account for all request actions against classified reports. This made it possible to develop some usage statistics and thereby devise a formula which could be applied for reproducing popular reports in quantity. The copies thus reproduced were placed on the shelf with the result that subsequent requests were handled more quickly, and the cost of reproduction was appreciably reduced.

As mentioned, this procedure applied only to classified reports which account for about 30% of the total requests received daily.

The procedure could not be applied to unclassified reports because no accountability was required, and inventory control of unclassified material was too costly in the manual procedures. Obviously, the automatic inventory control which would be possible with EDPS constituted an important advantage. But to realize this advantage it was necessary to bring the remaining reports under inventory control.

How was this accomplished? ASTIA had this tremendous volume of reports on hand in various unknown quantities. A master inventory punch card was designed by ASTIA, printed in quantity and prepunched by Remington Rand. Physical inventory of holdings was scheduled during week-ends, and absence or presence of reports with number of copies on hand was recorded on the master inventory cards. Immediately following the physical inventory, the information was key punched, and subsequent request actions were recorded by updating the inventory master.

As a result, inventory levels and usage statistics are maintained in the document record which also contains the validation information. Thus, whenever a document is validated for release to a requester, the inventory information is used to determine the correct action to be taken to fill the request. At the same time, inventory and usage statistics are updated. If reproduction is required, the usage statistics are applied to the optimum re-order formula, the number of copies to be reproduced is determined and a punch card reproduction order is produced. This is analogous to a back order to stock in a conventional supply system. The fact that a reproduction order has been issued is recorded by showing the quantity on order. When the reproduction has been completed, the number of valid requests held against it is deducted and the inventory is updated.

Finally, the physical mass of stock copies on hand can now be effectively controlled by the EDPS. All inventory records are periodically screened by the computer in terms of usage since date of announcement by ASTIA. Reports which reflect very littley or no activity for a given period of time are automatically listed for destruction. The personnel who are responsibile for the maintenance of stock need only remove the listed reports making certain that they verify the quantity of copies of classified reports and certify the list as accurate.

# Chapter 4

## Human Aspects of ADPS
by S. Edward Pope

This review covers some of the human aspects and related operational problems of the installation of an Automatic Data Processing System (ADPS) by the Armed Services Technical Information Agency (ASTIA). The full impact of the total transition will not be apparent for some time, due to the incremental phasing-in of the ADPS installation and operation. Only the initial phase (document request processing and document inventory management) has been started (this began 15 February 1960); phase II (application to information control and retrieval) was scheduled to begin operation 1 July 1960 and has been rescheduled for 1 August 1960.

Psychological and emotional reaction of ASTIA personnel related to ADPS installation seemed to range between blind unquestioned acceptance of the system with the implication that it is capable of almost any feat, and blind unquestioned rejection of the system with the implication that it will never work satisfactorily. There was some indication when operation was originally begun that most people seemed to lean toward one or the other of the above extremes, with relatively few people demonstrating an equatorial feeling. However, there now seems to be a more general merging of these views demonstrating more pronounced willingness to critically examine, understand, and accept realistic capabilities and limitations of the ADPS.

Some of the other human aspects considered here are those related to:

a. Customer Relations

b. Impact on work life and habits of ASTIA associates

c. Future Effects

d. Conclusions

CUSTOMER RELATIONS

Initial operation of ADPS is revealing a need for great emphasis on an overall Customer Relations and Education Program to apprise the R&D organizations on the critical aspects of the parts of system with which they are directly concerned. It also reveals a requirement for improving our capability in day-to-day service contacts with the R&D community. Some of the rather routine operations into which the human factor has thrown a kink are illustrated by the following:

The Document-Request Form (ASTIA Form-1) was carefully developed to include pre-punching by ASTIA of the address code that distinguishes a particular organizational entity address from all other addresses in ASTIA files. Reserving enough space on the form for a printed address, in an area not susceptible to mutilation by punches, had posed a problem. It was felt that pre-punching the address code would adequately insure machine identification of the requesting activity upon subsequent return of the request form to ASTIA.

This feature worked fine, except for the use of the pre-punched blank forms by recipients in a sort of musical chairs game. Apparently some activities have borrowed the forms from other organizations after their own supply was exhausted; in some cases military agencies have passed their forms to some of their contractors. When the exchanged forms are sent to ASTIA, the document requested (if validated) is automatically sent the organization identified by the pre-punched code (probably to their surprise). We are now reviewing a format for possible change to provide for this and other improvements. A fair share of problems has also developed from the familiar but undesirable stapling and mutilation of punch-card request forms.

Some adjustments are required by the recipients of ASTIA service. This is required primarily by the change from a multi-copy uniset document-request form to a single copy punch-card document request form, (sample attached), and the use of a new type (punch-card) shipping notice.

## IMPACT ON WORK LIFE AND HABITS

From an overall standpoint, almost everyone of the 360 people assigned to positions in ASTIA has been or will be effected, in varying degree, by the installation of the ADPS. Some ASTIA associates will be directly affected in a conclusive and decisive manner by abolition or radical changes in their positions; other associates are affected less radically such as by changes in procedure, or in some instances are affected only indirectly. In between these two groups is the large group of individuals who will be substantially affected by the changes accompanying the automation of ASTIA.

### Temporary Effects

Some of the immediately felt changes in connection with ADPS were temporary in nature. Among these were the effect on the various associates who felt pressures resulting from the necessarily closely scheduled preparatory work for the ADPS installation and operation. This included: file preparation, conversion, and cleanup; preparation of the ASTIA Thesaurus; the close checking of large masses of items concurrently and immediately after the operation was started. These are all outlined in the other presentations.

In many instances the work was related to, but in addition to, regularly assigned duties; a great deal was on regularly scheduled overtime over an extended period, and in many cases involved special assignments and temporary detail of personnel to different positions.

The first two resulted in rather normal human reactions of physical and mental weariness, irritability, and general lowering of personal efficiency, in some cases. This would have been much less acute with a longer period scheduled for preliminary and preparatory work, or compensatory increase in personnel resources to the Agency, or an optimum combination of both.

The special assignments in many instances resulted in an interesting and enlightening experience for the associates involved. Human adjustments by associates have been required in relation to the use of numbers on a much larger basis and by modification in the techniques involved in the use of numbers. A change requiring human adjustment in working with numbers is required to match rejected document requests with the ADPS explanatory notice and a mailing address label which involve matching address code numbers on a large volume basis (sample attached). Sometimes the document number and address code number get confused. This contrasts with the previous procedure involving the use of the address for any matching under the manual system.

Another example is illustrated by document requests coming off ADP machine runs in descending numerical order, rather than previous conventional ascending order. In previous similar situations, other personnel have found this arrangement equally satisfactory, after adjusting to such a procedure, and no doubt ASTIA associates will develop a similar response.

## Major Permanent Effects

A high order of human adjustment is required by thirty of the associates in ASTIA who have been affected in a most decisive and conclusive manner by functional and procedural changes under which their positions have been phased-out (or will be by 30 June 1960). Four of these people have been, or are in the process of being retired; three having elected to take advantage of the provisions under Civil Service rules for involuntary retirement. The remainder of the above mentioned thirty employees have been, or are expected to be placed in other positions within ASTIA. This is being accomplished through attrition. A few of these employees have been transferred to the Data Processing Branch -- primarily clerks and clerk-typists who will be required to learn new skills. So far, approximately twelve other employees have elected to transfer into the ADPS operating positions because they considered it to be a good personal opportunity. Some of these were selected from Data Processing aptitude tests and trained in new skills (programming, equipment operation, key punching, etc.)

Adjustment by many other ASTIA associates, in most cases, is less drastic than some of the above, but quite substantial. This includes adjustment relating to a decrease of personnel assigned to certain functions The transition to an operational decision-by-exception philosophy results from capability of ADPS to handle routine request processing. This will place a premium on human capacity for consideration and decisions required by the large volume of remaining exceptions.

Some adjustments are easier. An example of this is that related to transition to the stiffer punch-card document-request form from the relative flimsy uniset form. The cards are much easier to manipulate and handle in the request filling and processing operations. Large improvement in persona efficiency has resulted from this in conjunction with the improved inventory control resulting from ADPS. Other substantial adjustment requirements will be raised in the near future in the Bibliography, Reference, and Identificat Service functions in which ASTIA associates have developed very good and ver rapid retrieval proficiency under our previous manual system. Under this sy control of the several hundred thousand scientific and technical reports was maintained through manually compiled card catalogs (sample attached).

Use of the card catalogs will be phased-out in favor of punch-card files containing similar information, but in a different format (sample attached). The punch-card files will facilitate mechanized checking and rearranging periodically to insure proper arrangement. This has not been feasible under the manual catalog system, because of the mass and volume involved - several million cards in the various catalogs. However, this too will require flexibility on the part of the people involved and adjustment to adapt personal techniques to working with the radically different punch-card files.

## FUTURE EFFECTS

As we progress more deeply into the planned automation program there undoubtedly will be great impact, both positive and negative, on remaining ASTIA associates, which can not be fully foreseen nor completely assessed at this time. Some of the possible effects are that more people will be displaced by Data Processing equipment and reshuffled into other positions. However, the adverse effects of this can be minimized, if new opportunities for improving ASTIA service, to the military Research and Development Community, are diligently sought. It is essential that a balanced perspective be retained by all concerned to prevent permitting ADPS to become the master rather than the servant.

One of the most interesting aspects related to future effects is long-range possibilities for a basic major revision in the Agency mission. ADPS may well open the door for evolution into a true information handling and evaluation organization instead of a document handling and evaluation agency, as at present. This is brought into the realm of practicability by the vastly increased capability for processing and assimilating information which ADP equipment provides. This is expected to become increasingly more feasible over a period of time as ADP equipment and systems are improved further for this purpose.

From a philosophical standpoint, there are considerable pluses, as well as minuses relative to Human Aspects. We shall not eliminate human aspects or human requirements for some time -- a long time in the future. In fact, this may well be the most critical requirement in the near future. The long-range implications for agency employees generally seem promising on the whole. Statistically, a relatively small number of employees have been adversely affected so far. However, it is somewhat difficult for an individual to be very philosophical about statistics or the long-range point of view, if he is one of the adverse statistics, or is even uncertain about his personal future.

## CONCLUSIONS

1. The importance of thorough, careful and deliberate consideration of all plans, schedules and decisions related to ADPS assessment, installation, and operations phase-out/in cannot be over emphasized.

2. The hazards of over acceptance and over estimation of ADPS capability, particularly to actually reduce personnel requirements, justify special attention and emphasis. This can produce gaps in an otherwise smooth running operation. The hazard can easily arise as a result of uninhibited natural enthusiasm for a new system on the part of management and other personnel.

3. Time for preliminary preparation, planning, training and indoctrinati[illegible] service testing of ADP operation, and procedures development and modification should be liberally scheduled to permit adjustment and flexibility when necess[illegible]

4. Preoccupation with details of planning and installing the direct ADP System should not be permitted to obscure the attention required by related affected operations to develop necessary modifications in a timely and orderly manner.

5. The training and indoctrination program for the ADPS installation should be planned to cover procedural and operational changes in related areas directly concerned, as well as in the direct ADP System.

6. Schedules for phase-out of manual functions, operation, and particularly personnel should be carefully examined to avoid premature phase-out.

DOCUMENT-REQUEST FORM (UNISET)

PREVIOUS EDITIONS OF THIS FORM ARE OBSOLETE. REPLACES ASTIA FORMS 2, 11 AND 12, APR 57, ASTIA FORM 3, MAR 58 AND ASTIA FORM 15, MAY 57, WHICH ARE OBSOLETE

☆ U. S. GOVERNMENT PRINTING OFFICE: 1959 — 505141

ASTIA CODE

1. ADDRESSEE

REQUESTING UNIT (Optional)

2. DOCUMENT REQUESTED FOR (Sign One Only)

a. CONTRACT NUMBER

SIGNATURE

TITLE

b. OFFICIAL GOVERNMENT USE

SIGNATURE

TITLE

DATE OF REQUEST

3. DOCUMENT INFORMATION

a. ASTIA NUMBER (List Only One)

| AD | ATI | TIP |
|---|---|---|

b. CLASSIFICATION

☐ CLASSIFIED ☐ UNCLASSIFIED ☐ UNKNOWN

c. CHECK TYPE COPY DESIRED (Single Copies Only)

☐ MICRO-CARD ☐ FULL SIZE ☐ MICROFILM

d. REPORT TITLE (Caution—Classify this form if title is classified)

REPORT PREPARED UNDER CONTRACT NUMBER

ORIGINATING AGENCY OF REPORT

ORIGINATING AGENCY REPORT NUMBER

4. MILITARY AGENCY CONTROL (Note: If Disapproved, Return to Requester)

| RELEASE OF THIS DOCUMENT IS | ORGANIZATION | SIGNATURE | SYMBOL |
|---|---|---|---|
| ☐ APPROVED ☐ DISAPPROVED | | | |

FOR ASTIA USE ONLY

STAMP AS APPLICABLE

POSTAL REGISTRY

DATE SHIPPED

| U | C | S | YES | NO | L | R |
|---|---|---|---|---|---|---|
| | | | | | | |

COPY NO. 1

ASTIA FORM 1 APR 59

DOCUMENT-REQUEST FORM (PUNCH-CARD)

ASTIA NUMBER (List Only One)

| AD | ATI | TIP |
|---|---|---|

CHECK TYPE COPY DESIRED (Single Copies Only)

☐ MICRO-CARD ☐ FULL SIZE ☐ MICROFILM

DOCUMENT REQUESTED FOR:

USER CODE ______________________
(Enter complete user code)

CONTRACT NO. ______________________

Q T F Catalog No.

P-31963

Printed in U.S.A. REMINGTON RAND

CLASSIFIED REPORT RECEIPT FORM

POSTAL REGISTRY NUMBER (Used for receipting)

DATE SHIPPED

Receipt is acknowledged of material listed above. (If Contractor, it is recognized that this material is classified and comes within the purview of the security agreement signed by this concern.)

| DATE RECEIVED | SIGNATURE |
|---|---|

CONTROL RECEIPT

SUBMIT REQUEST TO:
ASTIA
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

User Code

FRONT

IDENTIFYING INFORMATION
(CAUTION: Classify this form if title classified)

ORIGINATING AGENCY OF REPORT

| ORIGINATING AGENCY REPORT NUMBER | DATE PUBLISHED |
|---|---|

REPORT PREPARED UNDER CONTRACT (OR MILITARY PROJECT) NO.

REPORT TITLE AND AUTHOR

1. If ASTIA Catalog No. not known, fill in identifying information.
2. For controlled documents: Enter ASTIA Catalog No. on face of this form and fill in identifying information. Submit request through controlling agency.
3. See current issue of ASTIA Technical Abstract Bulletin for detailed requisitioning instructions.

P-31963A

MILITARY AGENCY CONTROL

| RELEASE OF THIS DOCUMENT IS ☐ APPROVED ☐ DISAPPROVED | ORGANIZATION |
|---|---|
| (Note: If Disapproved, Return to Requester) SIGNATURE | SYMBOL |

ASTIA Form 1 Jan 60 Supersedes all previous issues of ASTIA request forms.

REVERSE

DOCUMENT SHIPPING NOTICE CARD

Contractor Notification

SHIPPING NOTICE

ASTIA, USAF
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

OFFICIAL BUSINESS

POSTAGE AND FEES PAID
DEPARTMENT OF THE AIR FORCE

**VIA OFFICE**

THIS IS TO NOTIFY YOU THAT THE CONFIDENTIAL REPORT HAS BEEN SHIPPED DIRECTLY TO THE CONTRACTOR AS INDICATED BY CODE AND AD NUMBER ON THE REVERSE OF THIS CARD

**CONTRACTOR**

THE CLASSIFIED DOCUMENT LISTED ON THE REVERSE, WHICH YOU REQUESTED, HAS BEEN SHIPPED TO YOUR SPONSORING MILITARY AGENCY FOR REVIEW. THIS PROCEDURE IS REQUIRED BY MILITARY REGULATIONS. ANY FURTHER INQUIRY REGARDING THIS MATTER SHOULD BE ADDRESSED TO THE MILITARY AGENCY

ASTIA FORM MAR 60 25

5840 AMP-T/2
THOMPSON RAMO WOOLDRIDGE, INC.
RAMO-WOOLDRIDGE DIVISION
ATTN: TECH INFO SERVICES
P. O. BOX 90534 AIRPORT STATION
LOS ANGELES 45, CALIFORNIA

FRONT

Z

7601 31366

ASTIA FIELD OF INTEREST REGISTER

FIELDS OF INTEREST

VIA

CODE

F

CATALOG NO.

Z 1 5840918

CL

DATE

MO. DAY YR.

USER CODE

Printed in U.S.A. REMINGTON RAND P-31744

REVERSE

DOCUMENT SHIPPING NOTICE

Military Agency Notification

SHIPPING NOTICE

ASTIA, USAF
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

OFFICIAL BUSINESS

POSTAGE AND FEES PAID
DEPARTMENT OF THE AIR FORCE

**VIA OFFICE**

THIS IS TO NOTIFY YOU THAT THE CONFIDENTIAL REPORT HAS BEEN SHIPPED DIRECTLY TO THE CONTRACTOR AS INDICATED BY CODE AND AD NUMBERS ON THE REVERSE OF THIS CARD

**CONTRACTOR**

THE CLASSIFIED DOCUMENT LISTED ON THE REVERSE, WHICH YOU REQUESTED, HAS BEEN SHIPPED TO YOUR SPONSORING MILITARY AGENCY FOR REVIEW. THIS PROCEDURE IS REQUIRED BY MILITARY REGULATIONS. ANY FURTHER INQUIRY REGARDING THIS MATTER SHOULD BE ADDRESSED TO THE MILITARY AGENCY

ASTIA FORM 25
MAR 60

5493-N-E/S
RESIDENT INSPECTOR OF NAVAL MATERIAL
GERMANTOWN
C/O PHILCO CORPORATION
4700 WISSAHICKON AVENUE
PHILADELPHIA 44, PENNSYLVANIA

REJECTED DOCUMENT-REQUEST

ASTIA NUMBER (List Only One) 600713 10 316752

AD 316 752 , ATI | TIP

CHECK TYPE COPY DESIRED (Single Copies Only)

☐ MICRO-CARD ☒ FULL SIZE ☐ MICROFILM

DOCUMENT REQUESTED FOR:

USER CODE 4135-903

(Enter complete user code)

CONTRACT NO. DA-36-039-sc-78121

Q T F Catalog No.

4135903 0

User Code

CLASSIFIED REPORT RECEIPT FORM

POSTAL REGISTRY NUMBER (Used for receipting)

DATE SHIPPED

Receipt is acknowledged of material listed above. (If Contractor, it is recognized that this material is classified and comes within the purview of the security agreement signed by this concern.)

DATE RECEIVED | SIGNATURE

CONTROL RECEIPT

SUBMIT REQUEST TO:
ASTIA
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

Printed in U.S.A. Registron Basic P-31963

SAMPLE ADDRESS LABEL

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

POSTAGE AND FEES PAID
DEPT OF THE AIR FORCE

OFFICIAL BUSINESS

4135A-I/S

AMERICAN ELECTRONIC LABORATORIES INC.
ATTN: L. SUPINA
121 NORTH 7TH STREET
PHILADELPHIA 6, PENNSYLVANIA

# EXPLANATORY REJECTION NOTICE

HEADQUARTERS
ARMED SERVICES TECHNICAL INFORMATION AGENCY
AIR RESEARCH AND DEVELOPMENT COMMAND
UNITED STATES AIR FORCE

Arlington Hall Station
Arlington 12, Virginia

Request for ASTIA Document:

0000316752CONF

Changed to

TO:

User Code 4135903 Via Code 7604

Document requested is furnished herewith ☐

Your request cannot be filled for the reason indicated below:

☐ Document is not in the ASTIA collection.
or
Cannot be identified from information furnished.
or
Document number cited has been cancelled.

☒ You are not currently established for service on contract cited.

☐ Document has higher security classification than authorized by FOIR.

☐ Document requested is not within FOIR.

☐ Document was announced by ASTIA as information only. Consult TAB.

☐ Document released only to US m[illegible]ary agencies.

☐ Document released only to US Government agencies.

☐ Document can only be released with prior approval of controlling agency. Direct request to:

☐ Secret reports are furnished only in full-size form to fill non-military requests.

☐ Document not in stock but being processed. Your request will be filled at the earliest possible date.

☐ Document is part of an older ASTIA collection. Your request will be processed as soon as releasability is determined.

SEE CURRENT ISSUE OF ASTIA TECHNICAL ABSTRACT BULLETIN
FOR DETAILED REQUISITIONING INSTRUCTIONS

ASTIA FORM DEC 59 15-2 REFUSAL NOTICE

## DOCUMENT CATALOG CARD

AD- 232 804 Div. 15, 25 UNCLASSIFIED
(16 Mar 60)

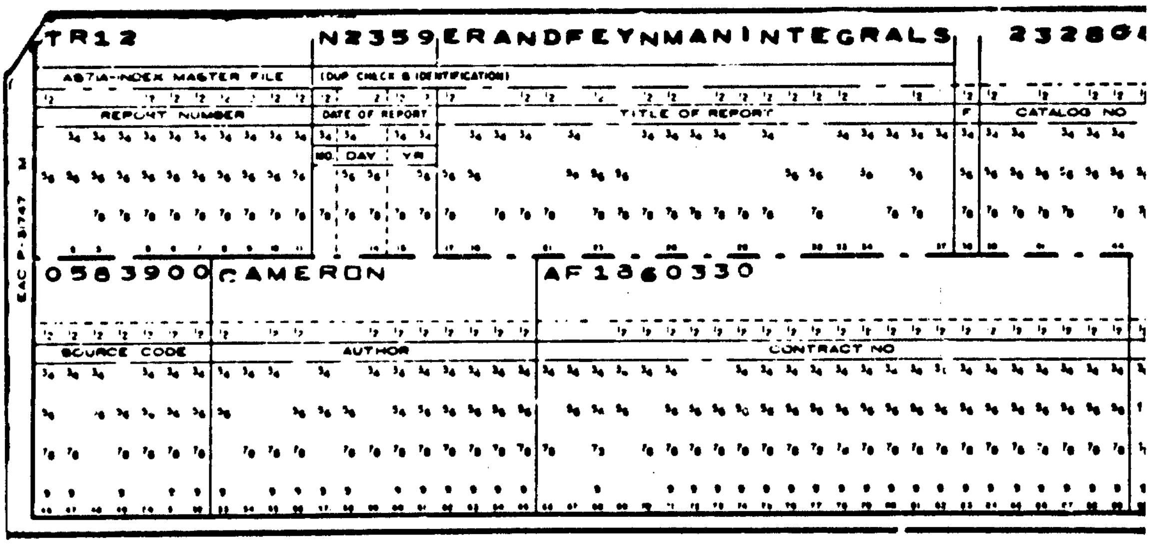
Minnesota U., Minneapolis.
A FAMILY OF INTEGRALS SERVING TO CONNECT THE WIENER AND FEYNMAN INTEGRALS, by R. H. Cameron. 23 Nov 59, 23p.
(Technical rept. no. 12)
(Contract AF 18(603)30)
(AFOSR TN 59-1272) Unclassified report

Descriptors: Integrals*; Functional analysis; Quantum mechanics*.

A sequential definition is given of a family of integrals depending on a variance parameter $\sigma^2$. If $\sigma^2$ is real this reduces, under reasonable conditions, to the ordinary Wiener integral. If $\sigma^2$ is pure imaginary, the integral can be considered as a Feynman integral. For certain analytic functionals, the Feynman integral is shown to be reduced to an ordinary Wiener integral. (Author)

UNCLASSIFIED

## DOCUMENT INDEX MASTER CARD



## Chapter 5

### Building an Information Retrieval System
by Herbert Rehbock

During our preliminary surveys, one of the areas in ASTIA which was found to be susceptible to electronic computer application was that of information retrieval. This is not to say that this was our first effort to mechanize information retrieval. In the late 1940's, the possibility of using punch-card techniques was explored. The idea was abandoned because of the inherent limitations of punch-card sorting and collation. For the next few years other possibilities were investigated and in the early 1950's a contract was initiated which resulted in the "Uniterm System of Coordinate Indexing." It was believed that this system could be used without machines. However, indexing and retrieval of a collection of the magnitude in excess of 200,000 reports defied mechanization and requires an Automatic Data Processing System. In ASTIA's case, it means an electronic computer with magnetic tape input.

It may well be wise to define what we mean by information retrieval since it conveys a different concept to each and every one of us. Information retrieval for ASTIA is the recovery of information from the technical reports collection when a request is received for a bibliography on a specific subject as well as broad areas of interest. Traditionally this operation has been performed by searching the subject heading card catalogs, a manual, time-consuming operation.

Although it was quickly realized that information retrieval is feasible, very little guidance could be obtained from either the published literature or from personnel conducting the preliminary feasibility study. ASTIA was on its own. During the early considerations we defined our objectives, which were to convert the AD document collection of approximately 200,000 reports to a machinable retrieval system, and to develop a system of indexing terms suitable for automation. A group was formed within the Document Processing Division and with the assistance of the Research Requirements Officer and the Automation Project Officer ideas were discussed on how the project could be approached. A multitude of ideas was brought forth, discussed, analyzed and discarded. One factor became apparent. ASTIA had at its disposal two systems of subject matter approach. One, the traditional subject headings, and two, the uniterm system. By 1959, the subject heading system numbered over 70,000 headings and they had served in a manual system with a proven degree of success. Uniterm (single word headings) had been assigned since 1953 to each report and those along with the subject headings were printed on the catalog cards. It was with automation in mind that the uniterm program was pursued. There was an apparent ease with which uniterms could be manipulated into machine language, and their retrieval accomplished by coordination or collation of related terms. An additional advantage was that the uniterms were selected from the actual document, thus imparting an authoritative character to the terminology as used by scientists and engineers.

As the realities of automation were faced in our discussion, it became obvious that neither the subject headings nor the uniterms answered the requirements. Both had much to offer, but there was much to be desired in the arrangement of the subject headings and uniterms, the clarification of semantics, the elimination of synonymous terminology, and the submergence of obsolete terms. What was really needed was a dictionary of machinable retrieval terms suitable for (a) assignment of retrieval terms to the documents in the collection, and (b) a tool for ASTIA customers for referance and look-up of terms used when requesting a bibliography on a specific subject matter. We decided to call the dictionary the "Thesaurus" and the retrieval terms "Descriptors".

Our first actual step towards this development of a thesaurus of scientific and technical descriptors was an effort to overhaul the "Subject Heading List". Principal headings were divorced from their subdivisions, and the list was reduced from 70,000 to 8,300 main headings, modifiers such as L-band and X-band were separated from the main headings. Subdivisions numbering 850 were reviewed and refined thereby eliminating 150 of these subdivisions as being no longer useful. This first step left us with a total of slightly more than 9,000 headings.

We had established a body of descriptors "authoritative" in character because they came initially from the reports processed by ASTIA. To define these descriptors and to give depth of retrieval to the Thesaurus, scope notes were developed showing what the descriptors included, what they were related to, or what they are limit ' to. This editorial exercise reduced synonyms, subordinated and submerged other terms, and permitted a further reduction in the number of descriptors from 9,000 to a little less than 7,000 units of information. A representative portion of the Thesaurus is shown in Appendix I.

Gaining a point of entry into 7,000 alphabetically arranged descriptors was the next problem which faced us. Certain accommodating tools had to be developed for use by both the subject analyst and the reference and retrieval operation. Check points and guidelines for quick, easy consultation were needed at both the input and output side of the storage and retrieval program. The organizational structure of the Scientific Analysis Branch seemed to provide a tailor-made answer. Here, our day by day processing of reports is subdivided into major areas of several organizational segments dealing in specific scientific disciplines. We decided to arrange the descriptors into related groups bringing such information elements as "hydrazine, borohydrides, liquid oxidizers", etc. together. Furthermore, it was decided that a descriptor should appear only in one group and in that group where the primary scientific or technical relationship is established. This task was accomplished by the chiefs of the scientific sections and other key personnel of the Branch who sorted and arranged the descriptors now in form of a deck of punch cards. Considerable discussion, reshuffling of punch cards and transferring of descriptors from one group to another occurred. Finally, the groups called schedules, were completed forming a distinctive display of related descriptors. Reanalysis of each group formulated a name for each schedule encompassing items such as: Mechanics, Space Propulsion, Pest Control and Inhibiting Agents. Schedules were arranged alphabetically and assigned sequential numbers. A total of 292 schedules developed, of which four are general in nature. A sample schedule is shown in Appendix II.

At last we had reached a point where use could be made of a long and tedious undertaking. Retrieval terms (descriptors) had to be assigned to

200,000 AD documents. Originally ASTIA had planned to develop the Thesaurus in house and to contract for the assignment of descriptors to be performed by a commercial organization. Two companies showed interest in the project but never submitted proposals or estimated cost. ASTIA found itself in the position of a necessary "in house" job which must be performed without an appreciable increase in personnel strength. An appreciation of the magnitude of this effort can be pictured by comparing the fact that the descriptor assignment must be completed in seven months, whereas the original analysis by means of subject headings and uniterms took seven years. Here was "data clean-up" to the nth power.

Before tackling the whole project, we spent some time on test assignment of descriptors to develop concepts and establish some systematic procedures for assignment. Several flaws were recognized and steps taken to overcome the deficiencies. The first one was the question on how to attain some degree of consistency in retrieval term assignment when 25 different people analyze reports. Has the analyst covered the subject matter, the equipment, methods, results, etc.? An analysis of reference questions in terms of the foregoing and other elements will surely assist in selecting the pertinent descriptors for retrieval purposes. An actual check list was developed and is as follows:

<u>CHECK LIST FOR ASSIGNMENT OF RETRIEVAL TERMS</u>

1. <u>Subject</u> (What was studied, investigated, tested, compiled, researched)
2. <u>How Was the Subject Treated?</u> (Analysis, tests, design, production, computations, theory, specifications, operations, processes)
3. <u>What Are the Physical Factors?</u> (Mechanical properties, physical properties, chemical effects, biological factors)

4. What Equipment or Method Was Used to Support the Research or Investigation? (Spectrum analyzers, oscilloscopes, Charpy V-Notch test equipment)

5. Where or Under What Environment Was the Research Accomplished? (Upper atmosphere, arctic, sub-surface, location if geographic or foreign)

6. Additional Qualifying Information (Open-ended Terms). (Project names, military symbols, trade names, Mark/Mod numbers, AN/numbers, etc.)

This check list was not expected to answer all of the possible conditions which may be encountered, but it does standardize assignment of retrieval terms.

In the check list we mentioned "open-ended terms". Here was another area which had to be considered before full scale assignment of retrieval terms would be attempted. There is in the military and commercial research a body of symbols, nomenclature, code and popular names which permit an entry point into a system as long as the information is recorded. Ground rules were established for format of the open-ended terms and the manner in which they were to be assigned. These terms were called open-ended terms, because they will not appear in the Thesaurus or in any of the schedules. They permit considerable freedom for establishing retrieval points where Thesaurus control is not considered essential. In many cases the open-ended terms are stated in different ways, but they pertain to the same equipment. This is not always apparent when the report is analyzed. For the computer the information is pulled together by assigning one code number to terms of synonomous meaning. Furthermore, this secondary system allows for the use of classified terms, thus keeping the Thesaurus an unclassified scientific and technical descriptor arrangement. With the addition of "open-ended terms" to the descriptors we began to refer to the total system as "retrieval terms".

Up to this point in time everything done for the retrieval system had been slanted towards alphabetical key punch operation. Since the key punching operation was to be performed by a contractor, serious consideration had to be given to cost. Here was an opportunity to prepare the descriptors. Such a system was developed by taking the alphabetic print-out of descriptors and assigning numerical codes in sequential order. A spread of 1000 between descriptors introduces flexibility into the system with a view towards future additions and revisions. (Appendix III)

Necessary print-outs showing schedule numbers, descriptor codes, and descriptors were provided and full-scale conversion of the 200,000 reports commenced. When a descriptor is assigned, the schedule number to which it belongs will be shown. The purpose of this is to prepare cumulative indexes for the Technical Abstract Bulletin (TAB). Although the first cumulative index which ASTIA will prepare by means of the computer will be arranged alphabetically by descriptors, it is feasible to issue future indexes grouped by schedules, thus showing related information in consolidated areas.

The 292 schedules and their related descriptors proved to be an adequate tool for assignment of retrieval terms, but something else was needed to guide a user of the Thesaurus to the schedule where the descriptors of importance to him are located. An analysis of the schedules showed a remarkable relationship of certain schedules. We grouped these related schedules and finalized an arrangement of 19 fields ranging from Aeronautics thru Space Technology. Each major field is given a name and below are listed the schedules related thereto. (Appendix IV)

Testing of a program as has been outlined is a highly desirable feature. However, for such an exercise sufficient material must be available for computer input. This was not the case, and therefore it was decided to test by rather crude methods the validity of descriptor assignment and retrieval. This was accomplished in two ways. ASTIA had prepared a bibliography on "Bio-Astronautics" in February 1958 which consists of approximately 1100 unclassified entries. We converted these catalog cards to descriptors so that tests by the computer can be run. Our second method was to have personnel of the Bibliography Branch and the Scientific Analysis Branch determine what descriptors they would actually assign or retrieve in order to fulfill a specific bibliographic request. The exercise was conducted under controlled conditions so that no information would pass from one group to the other as far as the test bibliographies were concerned. It was gratifying to note the excellent correlation of assignment and retrieval. We were, so to speak, transmitting on the same frequency.

Actual operation of the computer for information retrieval is planned for 1 July 1960. To this end, ASTIA is presently engaged in developing the program. At present we contemplate an initial search by an average of six descriptors followed by several sub-routines.

There remains the major task of getting the Thesaurus into print. By February 1960 the Thesaurus had been edited for the third time and we froze the design. Key punching was started, accompanied by constant verification and editing of punch cards. Our programmers have developed a program for the computer so that the high-speed printer will present us with a master copy for photo-reproduction and subsequent quantity printing Layout of the Thesaurus will be first by the 19 fields and 292 schedules, now called groups, to be followed by the descriptors arranged alphabetical and accompanied by their scope notes and related terms. Introduction and Foreword are in preparation, and the target date for publication is May 1960.

In summarizing the project I would like to restate the primary objectives of ASTIA in building an information retrieval system suitable for computer application.

1. Reduce the time required to prepare bibliographic and reference inquiries.

2. Improve quality of information services.

3. Computer prepared cumulative indexes for the Technical Abstract Bulletin (TAB).

4. Eliminate catalog card printing and maintenance.

5. Present ASTIA customers with a reference tool to our collection by means of the "ASTIA Thesaurus of Descriptors".

Audiometers
(Medical Equipment)
Also See: Hearing

Audiometry
(Medical Procedures)
Also See: Hearing

Auditory Acuity
(Physiology)

Auditory After-images
(Physiology)
Also See: Auditory Illusions

Auditory Illusions
(Physiology)
Incl: Agi
Audio-gyral Illusions
Also See: Auditory After-images
Psychoacoustics

Auditory Perception
(Physiology)
Also See: Pitch Discrimination

Auditory Signals
(Sound and Acoustics)
Also See: Sonar Signals
Sound Generators

Auditory Thresholds
(Physiology)
Incl: Binaural Thresholds
Monaural Thresholds
Also See: Deafness

Augers use Earth Augers; Ice Augers

Aureomycin
(Drugs and Biologicals)
Also See: Antibiotics

Aurorae
(Astronomy)
Incl: Australis
Borealis

Aurum use Gold

Austenite
(Metallurgy (General))
Also See: Iron
Phase Transition
Steel
Transformations

## Descriptors Groups (Schedules)

1 Accidents

17000 Accidents
150000 Air Sea Rescues
607000 Aviation Accidents
1766000 Disasters
1835000 Drowning
1851000 Dust Explosions
2788000 Hazards
3529000 Low Visibility Driving
3960000 Motor Accidents
5514000 Sea Rescues
6073000 Survival

2 Acoustic Detection

41500 Acoustic Detectors
50000 Acoustic Ranges
1862000 Early Warning Sonar
2984000 Hydrographic Sonar
2991000 Hydrophones
5483000 Scanning Sonar
5748000 Sonar
5749000 Sonar Beacons
5750000 Sonar Charts
5750500 Sonar Domes
5751000 Sonar Equipment
5754000 Sonar Projectors
5756000 Sonar Receivers
5759000 Sonar Targets
5762000 Sound Bearing Finders
5766000 Sound Ranging
6604000 Underwater Sound Equipment
6605000 Underwater Sound Generators

3 Acyclic Compounds

35000 Acetylenes
248000 Aliphatic Compounds
267000 Allenes
918000 Bromocarbons
947000 Butadienes
948000 Butenes
952000 Butanes
1038000 Carbon Tetrachloride
1190000 Chlorocarbons
1195000 Chloroprenes
1610000 Decenes
2090000 Ethanes
2101000 Ethylenes
2340000 Fluorocarbons
2761000 Halocarbons
2855000 Heptances
2862000 Heptanes

## Descriptors and Code Numbers

| | | |
|---|---|---|
| 93 | 1000 | Abaca Fibers |
| 23 | 1500 | Abdomen |
| 6 | 2000 | Ablation |
| 223 | 3000 | Abnormal Psychology |
| 217 | 4000 | Abrasion |
| 141 | 5000 | Abrasive Blasting Machines |
| 56 | 6000 | Abrasive Coatings |
| 217 | 7000 | Abrasives |
| 187 | 8000 | Absorption |
| 71 | 9000 | Abstracting |
| 73 | 10000 | Abundance |
| 204 | 11000 | Acaricides |
| 148 | 12000 | Acceleration |
| 268 | 13000 | Acceleration Tolerance |
| 200 | 14000 | Accelerators |
| 11 | 15000 | Accelerometers |
| 223 | 16000 | Acceptability |
| 1 | 17000 | Accidents |
| 214 | 19000 | Acepleiadylene |
| 14 | 20000 | Acetaldehydes |
| 88 | 21000 | Acetals |
| 17 | 22000 | Acetamides |
| 87 | 23000 | Acetates |
| 189 | 24000 | Acetic Acids |
| 24 | 25000 | Acetic Anhydrides |
| 31 | 26000 | Acetobacter |
| 19 | 27000 | Acetolactic Acid |
| 135 | 28000 | Acetones |
| 193 | 29000 | Acetonitriles |
| 135 | 30000 | Acetophenones |
| 190 | 31000 | Acetyl Radicals |
| 53 | 32000 | Acetylation |
| 15 | 33000 | Acetylcholines |
| 50 | 34000 | Acetylene Derivatives |
| 3 | 35000 | Acetylenes |
| 203 | 36000 | Achievement Tests |
| 201 | 37000 | Achlorhydria |
| 207 | 38000 | Acid Base Equilibrium |
| 201 | 39000 | Acidosis |
| 106 | 40000 | Acids |
| 78 | 41000 | Acoustic Decoys |
| 2 | 41500 | Acoustic Detectors |
| 249 | 42000 | Acoustic Filters |
| 100 | 43000 | Acoustic Fuzes |
| 249 | 44000 | Acoustic Horns |
| 249 | 45000 | Acoustic Images |
| 249 | 46000 | Acoustic Impedance |
| 249 | 47000 | Acoustic [illegible] eulation |
| 165 | 48000 | Acoustic Mines |
| 100 | 49000 | Acoustic Proximity Devices |
| 2 | 50000 | Acoustic Ranges |
| 279 | 51000 | Acoustic Torpedoes |
| 274 | 52000 | Acoustic Waveguides |
| 249 | 53000 | Acoustics |

## AERONAUTICS

(5) Aerodynamic Configurations
(6) Aerodynamics
(8) Aircraft Accessories
(9) Aircraft Categories
(10) Aircraft Control Equipment
(11) Aircraft Instruments
(12) Aircraft Structures
(33) Balloons
(41) Bombers
(63) Control Surfaces
(94) Fighters
(114) Guided Missiles or Missile Related
(134) Jettisonable Equipment
(137) Landing Gear
(198) Parachutes
(218) Propellers
(235) Research Planes
(254) Special Aircraft Features
(255) Special Purpose Planes

## BIO-SCIENCE

(21) Amphibians and Reptiles
(36) Biochemistry
(37) Biology
(38) Birds
(44) Carbohydrates
(86) Enzymes
(98) Food
(132) Insects
(139) Lipids
(143) Mammals
(144) Marine Biology
(162) Microorganisms
(169) Mycology
(199) Parasites
(210) Plants (Botany)
(221) Proteins
(222) Protozoa
(284) Vitamins

## CHEMISTRY

(3) Acyclic Compounds
(13) Alcohols
(14) Aldehydes
(15) Alkaloids
(17) Amides

## Chapter 6

### Impact of Automation on the Organizational Structure of ASTIA
by H. W. Miles

The preceding papers have discussed primarily what the computer is doing for ASTIA. The following and final paper will describe our plans for the future. The purpose of this paper is not to continue to discuss what the computer can do for us but, perhaps more importantly what the computer is doing to us.

Organizationally, Automatic Data Processing was first assigned to the Management Division as the Data Processing Branch. This Division also had the responsibility for management analysis, the programming and planning function, management engineering, financial management and statistical services. The Data Processing Branch was established in July 1959, approximately seven months before the computer was installed and operational. Manning of this branch, which now consists of 24 persons, began immediately. As mentioned before, the manning of this branch was accomplished primarily by re assignment of personnel within ASTIA. A "phase out" study, completed in Octo 1959, identified positions associated with automation that would no longer be necessary. The "phase out" study struck horizontally across the organization In the Document Processing Division, elimination of the Catalog Maintenance Section was planned and the descriptive cataloging activities are being streamlined. The entire Verification Branch of the Customer Service Divisior was affected and the Identification Section of the Search Branch is scheduled for reduction as soon as the conversion from a card configuration to a tape configuration occurs in July 1960. The Inventory Control Unit in the Storage Section has also been eliminated. At this moment the Reproduction Division Organization remains intact; however, ADP has dramatically changed the character of the functions performed by this division. Indexes to the Technical Abstract Bulletin will now be prepared automatically by the computer and it is planned that the text of the bulletin will be prepared by the use of punched paper tape in the Copy Preparation Section.

Ironically, gradual "phase out" of personnel in these functions and reassignment of personnel into the Data Processing Branch occurred at a time when the Agency was directed to reduce the number of personnel on board as of 30 June 1960 because of budgetary and personnel limitations. Fortunately, ASTIA's Management Improvement Program was beginning to pay dividends through improved production records achieved through improved methods and procedures. The impact of data clean-up and file conversion in anticipation of automation has resulted in increases in productivity rates in those functions not being automated. In other words, the management of ASTIA attempted to get its house in order concurrent with making plans for automation of the Agency. So successfully have the improvements been accomplished that the service time objectives expressed in our feasibility study (September 1958) were almost achieved prior to automation.

The Agency has five Regional Offices located at New York City, New York; Dayton, Ohio; Los Angeles, California; San Francisco, California; and Washington. These Regional Offices now provide decentralized reference service to the Research and Development Community authorized to use ASTIA's services in the above locales. The reference service, available to visitors to these Regional Offices, under the manual system did not contain the latest scientific information available in the Agency. This is because the reference service was based on card catalogs. A document cataloged today would not be covered in reference work until it had gone through all the processing which is required to put catalog cards in the file.

By contrast, the reference service provided by automation will by-pass card catalogs. As a result, reports will be represented in reference work within two weeks after they are cataloged, as opposed to a lag from four to six months in the manual system. Elimination of card catalogs removes the source from Regional Offices upon which scientists could go and obtain immediate reference service. Consequently, automation has had the effect of centralizing our reference service that was formerly available on a decentralized basis. Furthermore, our customers will be provided with cumulative, quarterly and annual indexes to our Technical Abstract Bulletin (being made possible by automation) which will enhance their reference capability in their own office.

An interaction that will occur as a result of mechanized information retrieval is the relationship of personnel in our Scientific Analysis Branch with personnel in our Search Branch. Mr. Rehbock discussed our information retrieval system which was designed for automatic subject control of our vast store of scientific information. The Scientific Analysis Branch concerns itself with the input side of the information in this system, and the Search Branch, the output side, with the scientist asking for the results. In the final analysis, the value of the organization to the military Research and Development Community will depend on the Agency's ability to produce the product or information needed by our scientific community. Some thought must be given to the integration of these two units, if not organizationally, at least the mental processes involved in how a scientist asks for information and the determination if the Agency has described that information in the same terminology, so that an inquiry to the computer will produce the desired results.

Another interaction that will be beneficial as a result of automation is the information obtained by the Data Processing Branch regarding the categories of information most heavily in demand with the activities of the Receiving and Selection function. ASTIA, currently by a joint Army, Navy, and Air Force regulation, receives ten copies of each scientific report produced by the military R&D Community. Obviously, the demand for these documents is not in the same ratio as they are received. The computer should be able to indicate the incident rate of requests for documents by categories, and personnel in the Receiving and Selection Function could request categories of documents having the greatest demand in larger quantities than ten. The responsibility for determining when the agency should request quantities of documents in excess of ten rests with the Chief of the Receiving and Selection Function who now, intuitively makes these decisions. This type of a decision under an automated operation should be made only after a correlation analysis has been made of the information provided by the computer. However, the Chief of the Receiving and Selection Function has a scientific background and not a statistical background; consequently, studies of this type should be made by qualified personnel in the Plans and Analysis Branch.

Another example of how the computer is affecting the organizational structure of the Agency is the change of responsibility for making the decision to pre-stock a document. In the manual system, someone in the Reproduction Division would decide that since there were multiple requests for a document they should reproduce X number more than currently requested. Formulae developed by personnel other than in the Reproduction Division will determine "How much" of an item is to be reproduced and "When" it is to be reproduced for the majority of requests. However, exceptions to the formulae--and there will be exceptions--will have to be individually evaluated. Who will make these evaluations? Personnel in the Reproduction Division or someone who can ascertain what has effected the inventory balance for particular documents and how these effects have taken place.

When the computer had been operational less than two months it was apparent that our day-to-day contact with our users is our most critical and pressing problem. One of our simple problems is the fact that our users still use a letter of transmittal for requesting documents. This, by itself, is no problem; but they invariably staple from five to ten requests (punched card forms) to the letter of transmittal. The punched card request form should state that customers are not to staple or mutilate them in any way. These forms are pre-punched with a user code for each individual customer. Some users exhaust their supply and borrow punched cards from other users. Since there is no name or address on these cards, because they are identified only by the established user code, you can imagine what happens when they borrow punched cards from someone else. These are but a few of the myriad problems that automation in its initial stages has brought to the Agency. At the present time, only personnel who are intimately familiar with the operations of the entire Agency and Automatic Data Processing intricacies are able to analyze a problem situation arising from the change in our procedures. To this extent, personnel from the Data Processing Branch have been involved with activities of our Customer Relations Staff. The Customer Relations Staff was established to handled customer service problems at the staff level instead of the every day-te-day operational type problems. At the present time, there is no one activity charged with the responsibility of the day-to-day problems that seem to be increasing daily.

As can be seen, Automatic Data Processing has had wide spread effects on the organizational pattern of ASTIA. Management will have to come to grips with these changes now and for some time to come. It can be said that the computer has cut across organizational lines, reducing layers of management and consolidating functions. The computer fits into an integrated system, the application of which includes automatic request validation, automatic inventory control, complete and accurate accountability for both classified and unclassified documents, mechanized index preparation for the Technical Abstract Bulletin, automatic duplication check for incoming documents, automatic identification of documents requested without reference to specific ASTIA catalog numbers, mechanized reference and bibliography service, and lastly, management information never before available.

In all the above applications, the source information obtained by our document processing personnel is used over and over. The basis of the existing structure is departmentalization. With Automatic Data Processing, these functions will all be performed in one place and the reason for the various activities is eliminated. One solution is to push the various activities together, reducing the number of sections and branches as functions are changed. Fewer layers of administration should result as there will be fewer sections and branches to be supervised and coordinated. This, I believe, will have a disturbing impact on some administrators since position and pay in Civil Service is based to some extent on supervision and responsibilities. Reorganization may well revolve around the Automatic Data Processing Branch. ADP in ASTIA has three functions: (1) Automation of the business type operation, (2) Information retrieval and, (3) Providing management type data. The computer will make possible the combination of problems that are alike for blanket decisions and the separation of problems that are different for individual decisions, which encourages the application of the most effort and analysis to the items which are important, and of the least to those items which are not so important.

The many changes occurring should enhance and increase the workload of management personnel through the development of operational programs for Command objectives, analysis and evaluation of operational data, and the pc trayal of progress toward these Command objectives. A quotation form James March and Herbert A. Simon's book, "Organizations", I believe, demonstrates this point.

> What determines the type of activity that members of an organization--and here we are particularly concerned with members at a relatively responsible level--engage in? We can cite two factors that affect the propensity of organization members to engage in an activity. First, the greater the explicit time pressure attached to an activity, the greater the propensity to engage in it. The stimulus of deadlines tends to direct attention to some tasks rather than others. Second, the greater the clarity of goals associated with an activity, the greater the propensity to engage in it. It is easier to attach rewards and penalties, internal as well as external, to completion of tasks with clear goals than to others.
>
> These propositions lead to a prediction that might be described as the "Gresham's Law" of planning: Daily routine drives out planning. Stated less cryptically, we predict that when an individual is faced both with highly programmed and highly unprogrammed tasks, the former tend to take precedence over the latter even in the absence of strong over-all time pressure.

If it were necessary to point to one single factor responsible for the success of the Agency in establishing an implementation date for conversio to Automatic Data Processing and meeting that schedule, it would be the pressure and explicit instructions received from top management. In ASTIA, systems analysis and the planning and controlling functions work to a larg extent through management processes located at the staff level and readily accessible to the Commander. Operational information flows through this management process for review and analysis and presentation to the Commander and his staff. Statistical information or new knowledge about our product and the consumer of our product will be one of the main benefits from our automation program. In its broader aspects, inventory control is really accomplished by consumer research. It might be said that consumer research is communication between the manufacturer and users and potential users of a product.

At this point, I would like to continue with another organization chart depicting ASTIA in relationship to its users. ASTIA is an Air Force organization under the management control of the Air Research and Development Command. Participants in the ASTIA program are the Army, Navy, Air Force and their contractors. ASTIA receives scientific and technical information from our users and disseminates this information to interested users having the same problems or working on research that may already have been completed by another organization. So, in effect, ASTIA is a tri-service organization for the Department of Defense, and to an extent is a mirror of all the research and development being conducted as reflected in the report literature required to be published as a result of this research. Information regarding the output of this effort matched with the demand for this information should not be of interest solely to the management of ASTIA, but to the Department of Defense. This is one instance in which I believe the top officials in ASTIA have been geared directly into the middle of Automatic Data Processing and are prepared and anxiously awaiting the type information the system is expected to produce.

The command element of ASTIA consists of the Commander and Director who is military and responsible to the Commander, Air Research and Development Command, for implementation of the approved program. The Deputy Director is a civilian who fulfills the responsibilities of the Director in his absence and acts as his principal assistant and advisor in the formulation of ASTIA policies, plans, and directives. The Executive is also military who is responsible for assuring that the policies of the organization as applied to the details of day-to-day operations are carried out.

The role of the personnel in the command element of ASTIA, which is primarily a civilian organisation, will be enhanced to the extent that factual data will be available for the first time on the patterns, habits and customs of the agency's users. This type of information will be of benefit not only to members of ASTIA but to each of the respective services and the Department of Defense. In addition, the Commander of ASTIA is a member of the Advisory Group for Aeronautical Research and Development, Documentation Committee. This organization exists to promote cooperation in research and development among NATO countries.

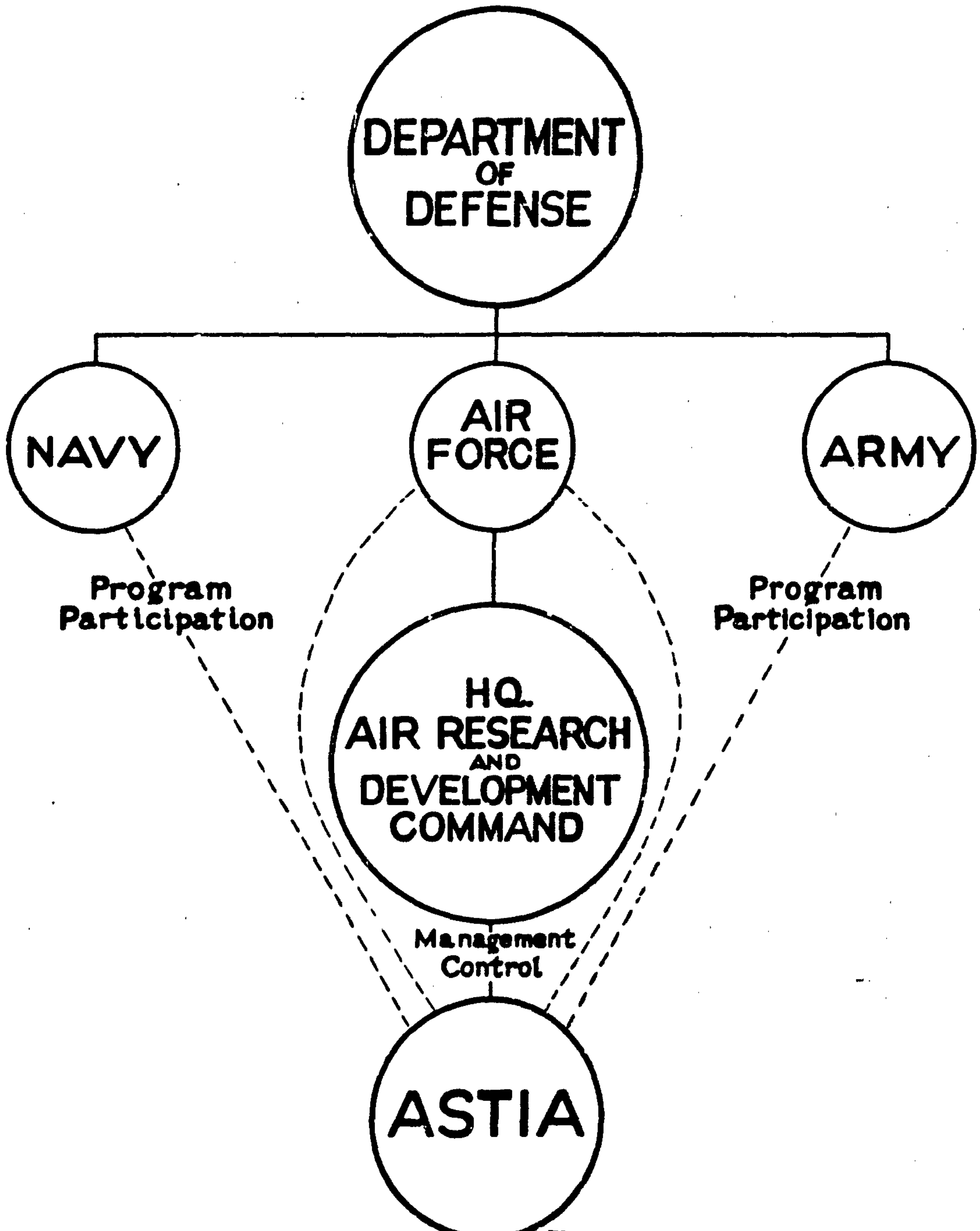
The role of the Commander and Director could conceivably be that of primary concern with external forces at a higher level. Decisions at Headquarters of the respective military agencies regarding contemplated research could be influenced by knowledge available in ASTIA on the state of the art, which could conceivably eliminate duplication and prevent unnecessary expenditure of funds. Effective information retrieval is a common problem of many government agencies and is a technique that

ASTIA plans to exploit to the fullest extent possible with its automated operations. Again the Commander and Director would be concerned with "value" analysis of the organization's impact on other U.S. and foreign scientific and technical information activities involved in similar functions.

The role of the Deputy Director and the Executive could be that of primary concern with the internal affairs of the agency and the improved effectiveness of the production force. Automatic Data Processing should permit optimal allocation of personnel resources and scheduling of productive work effort.

Chart No

# ASTIA TRI-SERVICE RELATIONSHIP



## Chapter 7

## THE ROAD AHEAD
## INTEGRATED DATA PROCESSING (IDP)
### 6 April 1960

The preceding papers have described the various aspects of ASTIA's program for EDP. My purpose is to tell you about our plans for the future. Colonel Hammond has pointed out that in the past ASTIA had no data processing equipment. Everything was done manually with an assist from mechanization in certain functional areas.

Because of this situation, our approach to EDP has been of the "Middle of the Road" concept. That is, we selected the functional areas in which the greatest gains could be quickly realized in terms of reduced processing time and increased capability; also, we chose to convert the processes "as is". We feel that we had ample justification for this because the systems analysis which went into planning for EDP had refined the existing procedures to a point where we were confident that everything we were doing needed to be done.

In our feasibility study we had contemplated using punched paper tape equipment in preparing reproducible copy for our Technical Abstract Bulletin (TAB) and catalog cards. However, the use of punched paper tape was not firmed up as part of the original plan because of the additional workload that this would impose on our planning and programming staff. Furthermore, at that time it simply represented an improvement in the method of preparing TAB and catalog card copy. It did not relate directly to EDP.

Shortly after our EDPS plan was approved and preparations had begun in earnest, it was decided to proceed with plans for introducing punched paper tape equipment into the copy preparation function. This decision was made because the EDP equipment which was finally selected had certain characteristics which could be capitalized on to advantage in terms of Integrated Data Processing.

As we moved into this area, it became apparent that the punched paper tape which was to be created initially in the course of preparing reproducible copy could be converted to magnetic tape if proper provisions were made for such conversion. The information thus recorded on magnetic tape could be printed out on the high speed printer as a result of the bibliography search. In addition, much of the information which is to be "captured" in punched paper tape can be machine converted into punched cards for the Index Master, Inventory Master and Information Retrieval files. In order to assure accuracy of the information, we are planning to use master edge punched cards for entering certain authoritative information into the TAB copy and thus into paper tape. Prospects such as these tend to make Management extremely eager in terms of expanding the applications of EDP.

The advantages of doing what has just been described are obvious. We can eliminate the duplicative tasks of key punching and verifying this same information. Here again we come face to face with the fact that there is more systems design and actual data processing programming to be done than our staff can handle all at once. Therefore, it is necessary to phase into this aspect of integrated data processing on a time available basis. We can't afford any slippage in reprogramming the punch card applications for magnetic tape. And we can't afford any slippage in programming other applications that were planned for the tape configuration.

We can determine what special coding to provide in the punched paper tape in order to permit the computer to handle format in the print out of bibliographies. In addition, we can prepare the program master tapes that will be needed in the copy preparation function.

At this point, perhaps I should digress a bit and explain why it is necessary to develop program master tapes for use with the punched paper tape equipment. The Technical Abstract Bulletin is an unclassified publication; yet it lists all documents cataloged by ASTIA, and some of these are classified Secret. Consequently, we must be sure that classified information is excluded from the TAB. On the other hand, bibliographies are geared to the requester's "need-to-know" and security clearance; hence, all the information concerning a particular document must be captured in punched paper tape for subsequent use in bibliographies.

This necessitates the typing of all information in order to record it in punched paper tape. However, by using a program master tape a by-product tape is produced which can be read by the machine at top speed to prepare the TAB reproducible copy. The program master tape inserts control codes into the by-product tape in such a way as to suppress the classified information while machine typing the TAB reproducible copy. No by-product tape will be punched in this latter process since the paper tape that is to be converted to magnetic tape was produced as a by-product tape at the time of typing the complete entry as described earlier.

There are a number of punched paper tape machines on the market. However, to our knowledge only two are presently capable of producing copy of reproducible quality. We tried both of these machines and selected the Synchro-Tape. We chose the Synchro-Tape simply because we felt it offered greater advantages for our particular application.

Earlier in this paper I mentioned we were confident that everything we were doing needed to be done. As we progressed in our planning for the punched paper tape operation we realized that we could dispense with catalog cards for all but about 2% of the documents cataloged. Our original plan for EDP had eliminated any need for card catalogs, but the bibliography service was based on stock-piling catalog cards to be manually withdrawn by the number after a machine search which produced a list of document numbers. The capability of machine printing bibliographies removed the remaining requirement for catalog cards for all except those documents in which special characters were used. In this instance, the term "special character" means any character that is not available on the high speed printer or the Synchro-Tapes. Some of the Synchro-Tapes will have selections of special characters such as mathematical symbols and others which are peculiar to the fields of chemistry and physics which are not available on the high speed printer. These Synchro-Tapes will be used for preparing TAB entries which contain such characters. The punched paper tapes for these entries will be retained and filed by the AD (ASTIA Document) number. These tapes will be used for printing bibliographic entries for the AD numbers involving this type of information. We estimate that this will be no more than about 10% of all bibliographic entries, although in specific instances an entire bibliography might have to be produced in this way. The 2% which I mentioned earlier are entries in which hand drawn characters such as sigmas, brackets and chemical diagrams are required. We will have to print catalog cards and stock-pile them for use in bibliographies which cite these particular AD's. Obviously, our Scientific Analysts can do much to minimize this requirement.

These vagaries will be coded into the bibliography search tape so that our program for searches can print out a reference to a typed entry or to a catalog card as the case may be. Thus, it will not be necessary for an operator to scan the entire bibliography in order to determine if there are any requirements for this type of special handling.

At present, our plan calls for implementation of the punched paper tape procedures on 1 July 1960. In order to achieve a reasonable capability for machine printed bibliographies in a relatively short time, we will selectively convert the cataloging information for documents that have been cited in bibliographies over the past several months. At the same time all of the day-to-day input will be processed for TAB reproducible copy beginning 1 July 1960. In this way we expect to attain a 60% capability for machine printing bibliographies by the end of the first year. And we expect to attain 98% capability by the end of the second year. Since we must continue to stock-pile catalog cards for 2% of the reports, 98% capability represents the maximum possible. Obviously, it would be silly to run a machine print routine beginning 1 July -- if we were lucky we might get one or two entries. Because of this we will run machine searches beginning 1 July and continue the present practice of pulling catalog cards by the number.

However, we plan to discontinue the reproduction of catalog cards on 30 June and file the punched paper tape by AD number for documents cataloged July through December 1960. These tapes will be used for recreating the bibliographic entries on the Synchro-Tapes as required during this period. We will use this method through December 1960. By 1 January 1961 we anticipate that there will be a sufficient number of AD's on magnetic tape to warrant the machine printing run.

Another plus factor in our use of the punched paper tape is that we can experiment with the computer in terms of information retrieval which goes beyond that based on descriptors assigned by our Scientific Analysts. There is much to be done in this field. In fact, a number of documentalists of national repute are interested in utilizing the potential which ASTIA is creating for experimentation in the field of autoindexing.

The introduction of punched paper tape equipment will make possible our initial excursion into integrated data processing. Our next objective involves a functional area of more far-reaching importance in terms of the number of personnel involved and the service time which we are able to attain. This functional area is reproduction of requested reports.

Each document cataloged by ASTIA is microfilmed as part of the input processing. This is done in order to have copies of all documents in a disaster storage area <u>and</u> to provide a means of reproducing reports as necessary in response to requests.

In our current operation, individual microfilm copies of the documents are filed by number. These must be manually selected, screened for type of reproduction to be used and spliced together according to the method of reproduction to be used. After reproduction the film must be broken down into individual document lengths, replaced in their individual cassettes and returned to the microfilm storage vault where the films are again filed by document number.

At this point it is in order to briefly explain why different methods of reproduction are used. One method, known as Xerox, is an electrostatic continuous process which does not require the use of sensitized paper. It operates at 20 feet per minute and is relatively inexpensive. Our reproduction of single copies of individual reports by the Xerox method costs about 2.5 cents per page. The other method of reproduction from microfilm is known as Airgraph. This, too, is a continuous process, but it is a photographic process which requires an exposure (or printing) step using sensitized paper and a developing, fixing and drying step. The second step operates at about 12 feet per minute. The Airgraph process costs about two times as much as the Xerox per finished page.

The Xerox process is satisfactory only for text and line copy. It is incapable (at present) of reproducing half tones and continuous tones. The Airgraph process, being photographic in nature, is not subject to these limitations. About 40% of documents in our collection contain half tones or continuous tones and must be reproduced by Airgraph.

Obviously, we could mechanize the selection of film and the reproduction process. However, the benefits which could be realized would be marginal in terms of the cost of systems design and development of the mechanization itself. In addition, we would still be saddled with two reproduction processes one of which costs twice as much per page as the other.

In the course of studying this problem, we have learned that at least two commercial concerns are exploring the use of video-magnetic tape techniques for recording and reproducing documentary information. Basically, a page is recorded on magnetic tape in a manner similar to the magnetic tape recording of a television program using a modified TV camera. In reproducing a page, the image is created on a device which is very similar to a television receiver. As you all know, such an image consists of lines -- this means that we could eliminate the Airgraph process from further consideration and concentrate on an electrostatic reproduction process.

Another commercial concern has already marketed a high speed "printer" known as the SC-5000. Actually this device, as far as we are concerned, is not a printer at all. It consists of a charactron fitted up with a Xerox processor for recording the images which are created on the screen of the cathode ray tube. With the Xerox part of the SC-5000 and some fairly simple circuitry we could have a means for reproducing our documents automatically and at high speed.

Furthermore, since the reproducible copy of each document would be in the form of magnetic tape, it should be relatively easy to design a system in which our computer would talk to the reproduction hardware and tell it which documents to reproduce for what order numbers. Obviously, the reproduction equipment would require its input in the form of punched cards so as to eliminate any waiting on the part of the computer or a sizable memory in the reproduction equipment to accomplish the same purpose. Concurrently with the reproduction, the computer could crank out the necessary shipping and accountability paper work which would reach the shipping operation in advance of the reproduced copies of the documents.

Perhaps this seems like fanciful "ivory tower" daydreaming. But bear in mind that we will fill requests for about a half million copies of reports this fiscal year; there are some 45 people involved in reproduction of requested reports and an additional 18 involved in the stock operation. In addition to the personnel, there is a little matter of about 10,000 square feet of space used for document reproduction and another 20,000 square feet of space used for document storage. Space and supplies are costly; personnel are very costly and are becoming more so as time goes by.

Annual costs for documents to fill requests break down approximately as follows:

| | |
|---|---|
| Supplies | $270,000 |
| Space (30,000 sq. ft. @ $1 per sq. ft. per yr. | 30,000 |
| Equipment Rental | 49,000 |
| Personnel (63 @ $4,000 pr yr.) | 252,000 |
| Total Direct Cost of Copies @ 500,000/yr. | $601,000 |

Compared to some of the large scale supply operations, this amount, i.e., $600,000, is peanuts. But it is approximately 25% of ASTIA's annual budget. As our workload increases, it will become an even greater portion of our budget. The point I am trying to make here is that making copies of reports available for filling requests is of such a magnitude that we can justify careful consideration of some pretty sophisticated hardware on an economic basis alone. When we consider the possibility of substantial improvements in servicing time and reduced unit costs which can be expected with advances in technology, we can see that a move to completely integrated data processing in the request processing area will undoubtedly be justified within the next few years.

Given this type of automation and integrated data processing, we can eliminate the inventory aspects of request processing. This will constitute a gain in terms of the complexity of our present data processing system which will in effect free the computer for other applications. This leads up to my final point which is: Systems analysis is a never ending task. As we have seen in this presentation, certain jobs that are essential today may not be essential tomorrow and some that are still essential tomorrow may not be essential the day after tomorrow.

Chapter 8

# THE "TAPE TYPEWRITER PLAN," A METHOD FOR COOPERATION IN DOCUMENTATION

Calvin N. Mooers

## INTRODUCTION

The "Tape Typewriter Plan" is a new program which I believe will provide a substantial advance in the technology of documentation. The immediate improvements will be of two kinds. First, the Plan provides a simple new means for cooperation between libraries to improve cataloging and to eliminate duplication of effort in cataloging. Second, the Plan provides a means for greater clerical efficiency in the cataloging operations now being done at each library.

The subsequent improvements in documentation due to the Plan will be even more important. The Plan will lead to inter-library cooperation on a large scale, providing many of the advantages – and none of the disadvantages – of a "universal documentation center."

Although the Tape Typewriter Plan is a cooperative plan, involving the exchange of material between libraries, the plan requires no compulsion and virtually no standardization of theory or method. The Plan is designed to provide maximum benefits to the participants with a minimum of contributory work. Participation in the Plan will follow from a library's determination of its own self interest.

The Plan has other features. The Plan is designed to make full use of modern electronic data processing equipment (electronic computers). At the same time, the Plan definitely does not require any participating library to have computing equipment of its own. For full participation in the Plan, all that is required is for each library to possess a "Tape Typewriter." This is a unit of only moderate cost.

One of the main purposes of the Plan is to permit the work done in cataloging documents at one library to be used for making catalog cards at other libraries. An unusual feature of the Plan is that it does not require a narrow standardization of cataloging format (arrangement of printing on a card). Each library can use its own format. Neither does the Plan require a standardization upon any single universal classification system nor descriptor categorization system of any kind. Libraries may use whatever classification system or subject heading system is most suited to their own operations.

While the Tape Typewriter Plan is very simple in basic concept, study will show that it has many subtleties and that it will have many far-reaching effects. I will now describe the Plan, piece by piece. Following this description, I will consider some of the consequences of the Plan.

## THE TAPE TYPEWRITER

The basic device, around which the whole Plan revolves, is a machine which I shall call the "tape typewriter." There are a number of commercially available tape typewriters. These are manufactured at various prices. Each make has various features of speed, convenience, and versatility of operation. All of these machines should be examined in the light of a specific library's problem before a particular machine is chosen for use.

Commercial tape typewriter devices are sold under the following names: *

| | |
|---|---|
| Flexowriter | (Friden Inc., San Leandro, California) |
| Olivetti | (Olivetti Corp. of America, New York) |
| Synchro-Tape | (Remington Rand Univac, Div., Sperry-Rand, Inc., New York) |

* See "The Computer Directory and Buyer's Guide, 1960" in Computers and Automation, Vol. 9, No. 6 (June 1960), and "All About Paper Tape," Datamation, Vol. 5, Nos. 3 and 4 (May-June and July-August, 1959).

| | |
|---|---|
| Teletype | (Teletype Corp., Chicago) |
| Telex | (Siemens & Halske, West Germany. U. S. representative: Siemens New York, Inc., New York) |
| Soroban typewriter controllers | (Soroban Engineering, Inc., Melbourne, Florida) |

The IBM "Cardatype" machine, although it produces punched cards instead of tape, should be studied also before making a choice of machine. There undoubtedly exist other useable machines, since my list is not intended to be exhaustive.

One of the important features of the Plan is that its operation does not depend upon universal standardization on any particular commercial make of machine. Each participating library can choose whatever kind of machine is found to be most available and suitable to that library.

For the benefit of those readers who are not familiar with tape typewriter machines, I now wish to explain them. A "tape typewriter" is basically a typewriter: it possesses a typewriter keyboard, and when one operates this keyboard, printed characters are made on a sheet of paper placed in the machine. In fact, most of the present commercial tape typewriters are built around a conventional electrically-operated typewriter mechanism.

In addition, the tape typewriter must also have two other features. It must have the ability to produce a perforated paper tape record of everything that is typed. It must also have a paper tape "reader" which can read the perforated tape and operate the typewriter in accordance with the perforations on the tape.

Consider first the production of the perforated paper tape. The machine must be able to make perforations or punches in a strip of paper tape each time one of the typewriter keys is punched. For each of the typewriter characters, a different pattern of punches is produced by the tape perforator.

Each pattern is produced by a row of punches placed in a single line crosswise to the paper tape. Some machines use a 5-location matrix for placing the punches in the line across the tape. In such a machine there are $2^5 = 32$ possible patterns of punches and blanks that can be recorded in any line on the tape. Such a tape can handle 32 different characters. As each typewriter character key is actuated, not only does this character print on the sheet of paper in the typewriter, but a corresponding pattern of punches is simultaneously perforated into the paper tape. After each character, the tape moves forward, with the result that as a line of text is typed on the sheet of paper, there is also produced a strip of perforated paper tape which records each step in the operation of the typewriter. In a fully capable tape typewriter, the letters of the alphabet, numerals, punctuation, "capitals," "carriage return and line advance," "tabulate," and "backspace" are all recorded by patterns of punches in the tape.

The 5-location tape perforating matrix is not the only kind used in current tape typewriters. Others use 6, 7, and 8-location matrices, with corresponding differences in the patterns of punches and blanks for recording the characters and greater versatility in the number of characters that can be handled.

The second additional feature that a tape typewriter must have is the capability to "read" perforated paper tape and to actuate either the typewriter keys or the tape perforator, or both, in accordance with the perforated tape record. Thus, a tape which has been produced as a consequence of typing one page of text, and which is then put back into the tape reader of the same machine (along with a new sheet of paper) will produce a tape controlled automatic typing of a page of text which is identical with the original typed page. It will also produce at the same time a new tape which is a copy of the first.

In summary, a tape typewriter has these elements. (1) It has a keyboard which is manually operated to produce both a printed page and a perforated tape at the same time. (2) It has a printing unit for printing characters on a page under the control of either the keyboard or of the strip of perforated tape in the tape reader. (3) It has a tape perforator for perforating paper tape in accordance with the keys pressed on the keyboard or in accordance with the tape passing through the tape reader. (4) It has a tape reader for reading a perforated tape and controlling the printing units and the tape perforator.

By full use of these elements, a number of very useful techniques are immediately made possible. Sections of tape can be pasted together to combine sections of text, and the resulting tape can be used to make a unified typed copy of the combined text. During the typing, a new tape can also be perforated.

One may let the tape typewriter print and perforate under the control of the tape in the reader until a blank area in the tape appears and the machine stops. Text can be inserted at the keyboard manually, and the tape can again be started through the reader in order to automatically print the remaining text as it is recorded on the rest of the tape. In this way, a new tape can be produced which contains the full corrected text exactly as printed.

By a slight variation upon this method, when the first tape is known to contain errors, the tape is allowed to run through the reader (producing typing as it goes) until just before the location of the error. The correction is then manually typed in. The tape in the reader is skipped ahead to the correct position after the error, and automatic typing is resumed. (Depending upon the particular machine used, there are differences in the details of this procedure.) While these operations are going on, a new and corrected tape is automatically produced.

From this description it is seen that a tape typewriter can be used to produce easily an error-free tape by making simple corrections in an error-containing tape. This error-free tape can be used thereafter to produce a perfect typed copy each time it is run through the machine. In other words, a tape containing the recorded text of a catalog card can be used to produce automatically as many error-free copies of the catalog card as may be desired. Single error-free copies of any difficult typed matter can also easily be produced in the same way. The implications of these capabilities to library typing procedures are readily apparent.

## TAPE TRANSLATION

We must now consider an important complication due to the fact that tape typewriters are not standardized devices. Not only do they differ between manufacturers, but machines made by the same manufacturer bearing the same model or type number will differ in important respects. The manufacturers are not to blame. Many of these variations are ordered by the customer when he buys his machine. Not only will a customer order various special symbols for some of the keys, but he may order a machine with tape having a 5, 6, 7, or 8-location punching and reading matrix. In addition, machines from different manufacturers may even use different patterns to represent the same printed character.

For this reason, the perforated paper tapes – as they come directly from a tape typewriter – are not freely interchangeable between machines. Each machine can use only its own particular tape and character pattern code. In order to go from one machine to another, with the second machine having a different tape and code, it is necessary to perform a simple translation and then to perforate a new paper tape which will be used to operate the second machine. Tape-to-tape translators of this kind are relatively simple devices. Laboratory translators have been built and have been used. When a real demand for translators arises, versatile models in commercial production can be expected to become available. Until then, electronic computers can perform this kind of translation.

With a tape-to-tape translator, the tapes prepared on any tape typewriter can be converted for use on any other tape typewriter. The only limitation is that the characters called for by the tape must be available on both machines. Through the use of such a tape translator, any tape typewriter can "talk" to any other tape typewriter irrespective of differing tape format or character pattern codes.

Only one tape translator will be needed to perform all the different kinds of translation for all the different tapes exchanged by a group of libraries. For each translation to be made, the tape translator is first adjusted for the tape format and the pattern codes of the source typewriter and the target typewriter. By use of a translator permitting such adjustments, one type of tape translator can serve for all kinds of tape translations.

## USE OF ELECTRONIC INFORMATION PROCESSING MACHINES

Tape translation is merely the simplest of the useful kinds of conversion procedures that can be performed on the tapes produced by a tape typewriter. Other more useful and more complicated conversions can be performed by means of electronic information processing machines (e.g., electronic computers). Such machines have a very high speed of operation, and their versatility is limited in most cases only by the imagination and inventiveness of the user. Time on these machines can be rented in intervals of a few minutes or hours. Such rental time is available on most computers at any of the commercial institutions which have them. Therefore, it is definitely not necessary for any library to purchase a computing machine in order to take part in the Tape Typewriter Plan.

Let us now consider several examples of the various useful conversions made possible by an information processing machine. Consider a set of tapes which have been prepared as a consequence of the typing of catalog cards, and which have been read into the memory of the information processing machine.

The machine can rearrange the sequence of entries making up each card, putting (for example) the author's name first. Then, inside its electronic and magnetic memory, it can also perform an alphabetization by author's name for a large collection of cards. When these alphabetization operations are completed, the machine can then perforate a new paper tape which will control a typewriter to type up cards in alphabetical sequence according to authors' names.

Any other rearrangement of content on the card (with or without sequence sorting) can just as easily be performed before making up a new tape.

Any part of the printed matter on the card (e.g., corporate authors, numbers, or abstracts) in the original card text can be omitted by the machine before making up a new tape.

New material (e.g., serial numbers, class numbers, etc.) can be added by the machine before making up a new tape.

The machine can change the "page format" before making up a new tape. Thus if the input tape was made for controlling the typing on a 5-by-3-inch card, the machine can compose new line lengths, set indentations and page margins for typing up an error-free standard $8\frac{1}{2}$-by-11-inch page. It can compose a single or double column format. Running heads and page numbers can be added. The output tape which is made will then control a tape typewriter which will type in this new format. Any kind of indentation, spacing, or other possibilities can be created automatically at high speed by the electronic information processing machine.

By various combinations of suppression of some parts of the text, and by format rearrangement, all sorts of new lists and compilations can be created from tapes originally punched for entirely different formats.

An information processing machine can just as easily take two tapes and merge their text contents into a single alphabetical or numerical sequence.

It can thus interfile items.

Conversely, the machine can select certain items from a list, ignoring the rest and make an output tape from only the selected items. Any criteria which can be numerically stated, or which involves any occurrence or combination of words, numbers, or classification symbols, can be used as the basis of selection. Therefore, any kind of "information retrieval selection" can be performed by the machine. The results can be perforated into a new paper tape for later printing.

In some cases, the output from a conversion operation by an information processing machine will be so voluminous that more rapid methods of printing would be preferred instead of the relatively slow (10 characters per second) tape typewriter. In this case, one of the electronic "high speed printers" will be employed. Some of these printers can print full lines of text (120 characters) at a rate of 10 lines per second, i.e., they can print a full single-spaced typewriter page in less than 6 seconds! This is a printing speed of 600 pages an hour which means that quite extensive lists and compilations can be printed in less than an hour of high speed printer time.

## USE OF AN INFORMATION PROCESSOR AS A TRANSLATOR

Because of its great versatility, an information processor can certainly convert from one typewriter tape code to another. It does this by means of a "table of conversion" which is stored in its electronic memory. For each code pattern on the input tape, it looks up the pattern in its table, finds the corresponding output code, and then perforates this second code on the output tape.

The use of conversion tables has many extensions, and these provide a challenging field of exploration. The input to the conversion table may be a number, a symbol, a word, or a group of words. The outputs can be just as varied. A few examples can be given: A table can be set up which takes words as inputs, and which produces a Uniterm number as the output. The inputs can be subject headings, and the outputs can be the corresponding numerical classification code symbols. The inputs may be words and phrases, and the outputs retrieval descriptors. The inputs may be code symbols in one classification system, and the output can be the one or more most closely approximating classification symbols or sets of descriptors in quite a different system.

It is my contention that whenever any two sets of "subject content indicators" (e.g., decimal classification numbers, subject headings, descriptors, etc.) are sufficiently well defined, it should be possible to set up a conversion table to go from the one set of indicators to the other. Of course, such a conversion, like any kind of language translation, will usually not be completely exact. Part of this inexactness is a result of the imperfect state of our knowledge of these systems and interrelationships. Another part results mainly from the inherent incompatibilities between the classification systems. In general, any given classification system does not make the same distinctions precisely as they are made in some other classification system. For this reason a single input symbol in one system may be represented by several output symbols in the second system. Other peculiarities may also appear, but a full discussion of these problems is inappropriate here.

The important thing to note here is that both the card or page format and the subject content indicators can be translated from one system to another. Because of this conversion capability, the Tape Typewriter Plan frees us from the need for prior agreement upon any universal subject indicating scheme or classification system. In this respect the Plan differs most significantly from all previous plans for wide-spread documentary cooperation, since these plans have all specifically required a prior universal agreement. In the history of documentation, such agreement has never occurred.

## UNION CATALOGS, BOOK CATALOGS, AND SELECTIVE BIBLIOGRAPHIES

Modern librarians would be far more interested in bibliographic tools such as book format catalogs, union catalogs, and easily prepared selective bibliographies if these tools were not so difficult and expensive to prepare and to maintain. For example, it is impracticable to maintain a book format catalog of the holdings of a library primarily because of the impracticality of making additions or frequent new editions. On the other hand, card catalogs, though clumsier to use, permit easy addition by simple insertion of cards. Union catalogs, with distribution of cards, run into their own problems because of the great clerical problems of filing and catalog maintainance. Thus such catalogs are a luxury. Printed selective bibliographies of several hundred or thousand titles, furnished to a customer at request, are another luxury that libraries have been unable to consider.

The Tape Typewriter Plan, because it is assisted by modern information processing equipment, places all of these goals within the range of achievement. I speak cautiously here, and only say "come within the range of achievement," because some present important library problems are so vast that they can still overwhelm the largest of our electronic equipments. However technological progress in equipment is so very rapid that even this limitation will not prevail for long. For another thing, many useful and significant library problems are not this vast, and so they can be handled now by present equipment.

## USE OF TAPE MEMORIES AND HIGH SPEED PRINTERS

Now I wish to discuss large-scale magnetic tape memories and high speed printers. These devices are adjuncts to electronic information processing machines. In a library plan, the tape memories will have a storage function similar to that of a card catalog. They will store bibliographic information in such a way that the information can be added, extracted, or rearranged at any time. Some of these memories, even at present, have relatively large capacities (e.g., enough capacity to store the contents of 100 drawers of catalog cards in one memory unit). Memories of this kind, at various locations, linked to the manipulating and translating capabilities of electronic information processors, and also linked via paper tapes to a multitude of tape typewriters in cooperating libraries, will present some truly intriguing possibilities for machine bibliographic communication.

These possibilities are even greater when the use of electronically actuated high-speed printers is taken into account. As mentioned previously, these devices can print a full page of text, typewriter size, single spaced, in less than six seconds per page. In one hour, 600 pages can be printed. Such printed copy can be used directly as a single copy or the printers can produce several carbon copies at the same time. The original copy can also be used for photo-offset reproduction for the printing of hundreds or thousands of copies. The cost of using high-speed printers is difficult to compute, but we can expect it to be within the range of a fraction of a dollar per page.

Using the large tape memories and high-speed printers available, it will be possible to furnish libraries with printed book-format catalogs of their holdings, or parts of their holdings, from the basic cataloging information stored on magnetic tape. Therefore, catalogs in book format will begin to compete with card catalogs, particularly when machine printing costs drop in comparison with costs of conventional card catalog maintenance. If and when this occurs, book format catalogs will become the day-to-day working tool.

At the site of the electronic information processing machine, the magnetic tape reels for the library will spend most of their time in storage. In fact, once the electronic machine has prepared a set of catalogs, it will thereafter be called upon to do library work only during short intervals on some periodic schedule. Thus any library will be obligated to support only a small fraction of the total charge for such machines. At other times the machines will be performing other jobs for other paying customers.

## COOPERATION IN DOCUMENTATION

By making use of the interfiling capabilities of the tape memories, the catalogs on tape of several libraries can be merged to give a combined or union reference tool. Again, the result can be printed out in book format, and copies can be furnished to the participating libraries.

The selective capabilities of information processing machines will permit a great many kinds of selection and arrangement of items from the master tapes to form selective bibliographies. With easy selection and printing of relatively long lists of citations (e.g., hundreds or thousands of items), selective bibliographies can be expected to become a much more useful and important tool than at present. In fact, it is possible that they will take their place as a major tool for information retrieval in the libraries that make full and modern use of the technique. In addition, a new technique known as "lattice indexing" now being developed by the author (which is related to techniques also being developed by others) indicates a new degree of retrieval power that can be given to machine-prepared selective lists and bibliographies.

Because of the practical value of knowing the geographical location of storage of documents, lists printed from this aspect can also easily be provided.

## THE PLAN IN OPERATION AT A LIBRARY

In the beginning, for each library, the Tape Typewriter Plan must justify itself solely on the basis of the improvements and economies that it can make in the clerical and typing aspects of library operations. In other words, the introduction of a tape typewriter should make it easier and cheaper to prepare multiple copies of catalog cards and to prepare error-free copies of lists and to prepare other kinds of error-free copy as needed in library operations. In fact, a tape typewriter should not be bought by a library unless it can pay for itself in these operations

In the next stage of the Plan, exchanges of paper tapes between one or more other libraries will be undertaken. Tapes for certain lists of citations and references, and tapes from certain categories of catalog cards will be discovered to be useful at two libraries. A trade can be arranged. A continuing "quid pro quo" exchange, with each library contributing something of value to the other, is easily set up. As such exchanges develop their usefulness, and other libraries wish to join and participate, the formation of tape exchange pools or joint collections will quite naturally begin.

At the individual library, as its collection of paper tapes accumulates (perhaps speeded by participation in a tape exchange pool), there will soon come the desire to do more things with the information collected on tapes. At this stage, the library will locate a nearby information processing machine with extra time available for rental, and will arrange to load its paper tapes into more concentrated magnetic tape storage. Now the library is ready to use the full gamut of electronic information processing capabilities: sorting, arranging, editing, translating, printing, and others. Catalogs and classified bibliographies in book form will be one of the major products to be asked for first from such a machine.

However – and this point is very important – one can only process information after somebody somewhere has prepared it on a tape typewriter. If catalog cards are ever made up on an ordinary typewriter, this labor at the keyboard has been wasted as far as machine processing, or inter-library cooperation via machine is concerned!

Therefore, the day is soon coming when it will be a "criminal waste" to put useful bibliographic information through a keyboard without at the same time producing some satisfactory kind of machine readable record such as a perforated paper tape, perforated card, or other medium.

THE PLAN IN OPERATION FOR A GROUP OF COOPERATING LIBRARIES

A group of libraries, cooperatively participating in the Tape Typewriter Plan, can use their force of numbers to gain a number of considerable advantages. By acting as a group, they can do better bargaining for electronic machine processing. They can also make better use of the machines. It is well known in the computing machine profession that the biggest (and costliest) machines are always the cheapest machines to use in terms of the number of machine operations performed per dollar spent. A group of libraries acting in concert can produce a continuing volume of work which will justify time on a large machine, and they can share the cost of providing "computing machine programs" for the various documentary tasks to be done. Furthermore, only by the formation of a cooperating group of a number of libraries can a union catalog operation be begun.

A typical group of cooperating libraries in a Tape Typewriter Plan will soon have available, through machine assistance, such a large and valuable pool of documentary material, that libraries outside the group will discover very compelling reasons to get a tape typewriter of their own and to join the pool and become participants. This should be encouraged. The rule of the game for full participation in any library pool should be "You furnish your tapes and we will furnish ours." However, special machine compilations and printing jobs for any particular library will have to be paid for by the library demanding such machine work.

As was mentioned earlier, it is not necessary for the participating libraries to standardize either on any specific commercial tape typewriter or on any specific catalog format. If differing classification or subject heading schemes are used, interconversion tables for these should be developed.

In order for larger libraries to handle their work loads, they will undoubtedly want to employ a number of tape typewriter units. These same libraries, in order to facilitate various tape exchanges, will also want their own tape translator units so they can communicate directly by tape with nearby libraries without having to call on computer assistance or on the use of outside tape translators. Smaller libraries will usually be able to arrange for tape translation (on some weekly payment basis) at the larger libraries. In this way the investment in equipment at the smaller libraries can be held to a minimum.

## WHY THIS PLAN IS SO UNUSUAL

The Tape Typewriter Plan is unusual in that it most ingeniously side-steps some of the intractable human and intellectual problems that have previously blocked progress in documentary cooperation. At the same time, it makes full use of the powerful capabilities inherent in all the mechanical devices employed.

All prior plans for large-scale documentary projects or for documentary cooperation have required (1) intellectual standardization, (2) format standardization, (3) disseminated labor, or (4) centralization.

Intellectual standardization has taken the form of a requirement for a common, universally agreed upon classification system (based on words, symbols, or numbers) which would be used by all the participants. Library history shows that such an agreement has never been possible.

Format standardization is necessary for any plan of catalog card exchange. However, even such a simple form of standardization as the use of uniform catalog card layout does not prevail among all libraries.

Catalog card exchange plans, of which there are many, are an example of "disseminated labor." This kind of group activity places a burden of labor – for filing the cards received – upon each local cooperator. This burden is not in proportion to the individual cooperator's selfish interest; instead the filing burden due to the flow of cards is in proportion to the size of the group. Therefore such plans, involving labor disseminated at the convenience of "others," all come to grief when the load increases beyond the level of local interest.

Many recent (and earlier) plans and proposals for large-scale bibliographic cooperation have been based upon setting up a unique bibliographic center. Reference questions, and all other information retrieval customer requests, would be processed by this one center before information could be provided to the customer. There have been many well-founded objections heatedly advanced against any such plan for unique centralization of bibliographic and documentary service. Except possibly in the Soviet Union, objections of this kind have prevented any widespread move in the direction of a universal information center

The Tape Typewriter Plan side-steps these four problems. Cooperation can take place without need to standardize on a classification system. (A detailed treatment of the considerations of this point will have to be given in another paper ) To a significantly useful extent, class symbolism can be converted from one classification system to another. Recording formats are easily changed by machine processing. Th an does not disseminate labor at the convenience of others. Each library does only as much work as it finds necessary in order to process its own documents only. Finally, the Tape Typewriter Plan leads to a loose confederation of libraries, each cooperating according to its own self interest, and using pooled magnetic tape storage at a variety of places – but certainly not at any one "universal documentation center."

The Tape Typewriter Plan is unusual also in the way that it recognizes and takes maximum advantage of the peculiar capabilities of electronic information processing machines. These machines can be considered to be high speed, completely accurate, clerical robots. Some of their greatest capabilities are in rearra 'ng data, looking up conversion lists of symbols, converting, storing, and printing. Properly used, these machines can do these things very cheaply. However, they do require a "machinable input" which the machine can read, i.e., perforated tape (or a set of tabulating cards used as a substitute for tape). The Plan provides this kind of machinable input.

Most people look upon a choice of a computer as a "final decision," and think that all subsequent work and thinking about machines must be based upon this one choice This view is erroneous, even though computer companies have tended to foster this viewpoint. The correct viewpoint is that computers can act as information processing machines. They can most capably perform high speed format and symbol translation between computers. Therefore if any group of libraries desires to make use of the services of a newer and faster computer, they can use the facilities of the new computer itself to convert from the magnetic tape format of the old computer to the new.

By using computing machines we can avoid some of the main problems that have plagued all projects of large scale cooperation in documentation.

## HOW TO GET STARTED IN THE PLAN

In order to get started in the Plan, it is not necessary to wait for anyone else to make decisions in any other library, nor to wait for further technological developments in mechanisms. You can start now. The machines needed for starting the Plan (the tape typewriters) have been commercially available for years. What you must do now is to study the capabilities of these machines in the light of your own library operations. It is likely that you will find that some of your clerical processes in document cataloging could be done more efficiently by use of tape typewriter machines. If this is so, you can begin immediately, saving the paper tapes produced as a by-product. Then, as soon as another neighboring library gets a tape-typewriter, you can begin an exchange of tapes in a "Two-Library Plan." More libraries will soon follow your lead. As soon as a large enough body of tapes is available, but only then, should an electronic information processing machine be called upon.

The beauty of the Tape Typewriter Plan is that you can start now at your library, on the basis of your own self interest. You don't need to wait for any kind of agreement on the part of others. From then on, whatever work you do – and preserve on tape – constitutes a resource for later more elaborate machine processing and for inter-library trading and cooperation, all to your benefit.

## APPENDIX

### THE USE OF NON-PRINTING CHARACTERS FOR FORMAT INDICATION ON ABSTRACTS TRANSCRIBED ON MACHINE MEDIA

A very serious problem arises in machine processing of abstracts and other documentary information. The problem is that it is difficult to instruct a machine how to distinguish between the title, the author or authors' names, the abstract, and so on. For this reason, it is desirable to have some sort of special "punctuation" to separate and to distinguish between these items. Such special punctuation should be easily interpreted by the computing machine. At the same time, such punctuation should not interfere with the ordinary print-out from the tape typewriter.

It is my proposal that a "non-printing character" be placed in the typewriter keyboard and in the tape code. This non-printing character will be recognized only by the computing machine. When this character is struck on the keyboard, no printing occurs, but the tape is punched with the tape code for the non-printing character (NPC). Conversely, when the NPC code is read from the tape, the typewriter does not print or space.

A single NPC will be adequate to handle all problems that will arise. This is because single and multiple uses of the one character can provide the required number of different meanings. Where the NPC is indicated by ∇, we can form: ∇, ∇∇, ∇∇∇, ∇∇∇∇, etc. Each of these is given a different meaning. By using a single "space" interpolated among the ∇'s, an even larger number of symbols can be formed, such as: ∇ ∇, ∇∇ ∇, ∇ ∇∇, ∇ ∇ ∇, etc. However, I believe it is preferable to avoid these latter combinations and not to assign a confusingly large number of NPC codes because there is a better way of handling these problems. I propose that for each transcription program, at each transcription center, a fixed sequence of descriptive items such as: document number, document title, authors' names, dates, etc., and abstract be set forth. This sequence will be called the "sequential format" of entries for each abstract. It is not necessary for each transcription center to use the same sequential format of entries. It is only necessary that the sequential format be determinate for any given piece of tape.

My suggestion for the assignment of interpretations to the NPC combinations is as follows:

| | |
|---|---|
| ∇∇∇∇ | Marker separating complete abstract or information entries. |
| ∇∇∇ | Non-printing "parenthesis" to enclose special comments. See below for further discussion. |
| ∇∇ | Marker to separate title from author, author from date, and in general to separate the various kinds of descriptive items within an abstract entry. |
| ∇ | Marker to separate multiple authors, multiple report numbers, and in general to keep content from running together within any one descriptive item. |

To use this method of format indication, a fixed sequence of descriptive items must be determined. For example, a typical "sequential format" for one run of tape might be:

1) document number
2) full title
3) authors
4) source agency or company
5) date
6) number of pages
7) abstract
8) classification

Transcription on this one tape must then follow this sequential format. Where any item of the format is omitted, there must be a "∇∇ ∇∇" with a space between the NPC's to indicate the omission. Where an unusual situation arises, this is taken care of by a comment inserted in this fashion: ∇∇∇ comment ∇∇∇. When this indication of "comment" is used, the computing machine can easily find the special comment and act appropriately.

The following example is given of the use of such a format and the NPC's.

> ∇∇∇∇ ADC 1012; ∇∇ Nuclear Fractionation. ∇∇ J. Jones, ∇ B. Smith; ∇∇ Reactors, Inc., and ∇ Univ. Texas, ∇∇ June 1959, ∇∇ ∇∇ No abstract, ∇∇∇ abstract is confidential ∇∇∇ ∇∇ Secret. ∇∇∇∇.

Such things as "carriage return," "indent," etc., have not been indicated, but their use is understood. Note that all the ordinary punctuation is retained.

From the example, the uses of the various symbols are quite evident. The only one needing further comment is ∇∇∇. The material enclosed by this non-printing parenthesis can be of great variety. One important use is to identify special material following the insertion, or to make special comments on deviations from the established sequential format. Symbols not in the regular type fonts can be indicated. By using ∇∇∇ in this way, all special cases not handled by ∇, ∇∇, or ∇∇∇, can be dealt with. Special words assigned conventional meanings can be enclosed in the ∇∇∇ bracket to indicate special functions, and these words can be looked up in a list or table in the computing machine to get the details of what actions are to be taken. Such a list can be of any length, and thus an arbitrarily large number of situations can be dealt with without the need for additional NPC codes or characters.

By the use of a special non-printing character, the computing machine processing of punched tape transcriptions of abstracts and other bibliographic information can be greatly facilitated. At the beginning of a tape run, the computing machine can be instructed in the sequential format of entries within each abstract entry. Thus various sequential formats can be interpreted without trouble by the machine. Consequently there will be no need for a universal standardized sequential format of transcription or print-out. This last consequence – the elimination of a need for prior standardization of format – is the most important result of this proposal.

## Chapter 9

## A Generalized Computer Method for Information Retrieval

by Mrs. Claire Schultz


### ABSTRACT

A generalized method is given for performing information retrieval by computer. The method can be applied to systems varying in type, size, manner of use, and computer equipment.

Systems with files of information-retrieval data or document references may use the method. Options allow many or few searches to be performed during each pass of the file through the computer. Alternative searches or subsearches may be specified, and the upper and lower bounds on the number of retrieval references may be set. The options can be used to define the characteristics of a particular system so that necessary staff and equipment capacities can be determined. Output forms include reference numbers, abstracts, microfilm, and so forth, depending on the output equipment chosen. In addition to retrieval information, the output may include system-use records, used to analyze and evaluate system efficiency automatically, and to compare objectively one system with another.

The first application of the method is by the Armed Forces Technical Information Agency. A block process diagram, detailed flow chart, and description of the ASTIA application are provided.

## SECTION 1

## DESCRIPTION OF METHOD

### 1-1. INFORMATION RETRIEVAL DEFINED

The method to be described is for performing information retrieval by computer. Since the literature does not differentiate information retrieval from other types of data processing, it may be worthwhile to do so here.

In what is generally termed "data processing," the computer manipulates specific units of input data to arrive at other specific pieces of data, such as the amount of a paycheck or the tangent of an angle.

In machine translation, the computer substitutes equivalents for one another, whether the unit size is a word, phrase, sentence, or otherwise. Again, the manipulation proceeds from specific to specific.

Artificial languages, such as those used in automatic computer programming, require a type of machine translation, and thus proceed from specific to specific. Expanding a pseudo-language (shorthand) to full length increases the output volume to more than the input, but is nonetheless a series of substitutions of one specific thing for another.

Information retrieval, however, arrives at a general class of information by examining specific units of input data for their interrelationships.

For example, to determine whether a document is pertinent to a particular question, the computer compares the document description with the units of information requested in the question. If they match, a question may further require that an additional unit or units of information be absent for the document to qualify as an "answer". After examining an entire file of document descriptions, one knows which documents meet the requirements of the question.

A question represents a class of information. The more units of information asked for in the question, the more exclusive the class. For example, a question which asks for documents mentioning dogs, cats, and rabbits is more exclusive than one that asks for documents mentioning cats and dogs.

In building a document-description file for information retrieval, all the classes of information for which there will be requests cannot be anticipated. For efficient retrieval, therefore, units of information in the input file must be small enough to allow great freedom in creating classes at the time of output. Specific input units must be planned so that the large generic units to which they bear a fixed relationship will be made explicit.

For example, if an input file contains information about apples, any document concerned with a specific kind of apple, such as Winesap or Mackintosh, should have the name of the specific apple in its description; but the description also should contain the word apples, so that if a question were received asking for information about the temperature at which to store apples, to perform a search, one would neither have to know the names of every type of apple, nor would one need to search for every specific kind of apple when one word, apples, would suffice.

1-2. EQUIPMENT

No particular configuration of equipment is presumed for the method described; the program, outlined in figure 1*, is feasible for almost any general-purpose computer system. As described below, it also is applicable to tabulating equipment, or even manually-sorted systems. Its efficiency declines, however, when desirable options (discussed later) are precluded because of equipment limitations. The program efficiency varies according to computer capacity to handle the necessary volume of searches and produce output in the desired format.

Small or little-used files could use the method with some kinds of tabulating equipment, if only one question were asked per reference-file pass. Most applications, however, will require answering more than one question per pass. An alternative for small files is to share computer time with other users.

To answer more than one question (and to perform system self-analysis and provide flexibility in optional features),

* Technical assistance in drawing figure 1 was given by Clayton Shepherd, Univac Engineering Center, Philadelphia.

the computer used should have (1) a memory which is large enough to store the file of questions being processed, and (2) memory access and arithmetic operations fast enough to allow the large number of necessary comparisons to be made in a practicable length of time. The speed of input-output devices may or may not limit efficiency, depending on the equipment used and the parameters of the particular retrieval problem concerned.

## 1-3. INPUT FILE

The method described in this paper presumes the existence of a well-ordered file, incorporating the above requirements. The input file might contain descriptions of documents, such as books, articles, patents, or musical scores, or it might contain descriptions of factual information, such as the properties of compounds, airplane seating arrangements, vital statistics, and so on.

The method assumes that the descriptions contained in the file will have been built systematically, and that the system is known to the persons who will interrogate the file. In many information-retrieval applications a thesaurus also may be used with the system described here.

## 1-4. OUTPUT

Using the method described, the output format can vary, depending on the needs of the system's clientele, the types of auxiliary files maintained to supplement the master file, and the configuration of retrieval-system equipment.

### 1-5. METHOD FLEXIBILITY

This information-retrieval method is extremely flexible because of the following features and program options.

### 1-6. SPECIFICITY

An outstanding feature of the method is that one can ask a specific question, but still provide compensation for a null answer through having asked too specific a question. For example, one might ask a question containing four keys, or descriptive units, of the system, which can be called A, B, C, and D. Referring to a previous example, these might represent cats, dogs, rabbits, and hamsters. If the file contains nothing about all four keys taken together, the system can be requested to give information about ABC, BCD, or CD, or any other grouping of keys. If keys are used that do not occur in the parent question, a new parent question is formed; thus, if one decided that the grouping of rabbits, hamsters, and guinea pigs might be interesting, the method described would consider this a new question. The machine program is completely flexible with respect to the number of keys a search may contain; however, the optimum number probably is three to six keys.

## 1-7. ALTERNATE SEARCHES AND SUBSEARCHES

The method is flexible as to the number of possible subsearches, using different key combinations in the parent question. It also is flexible as to the number of questions that can be asked simultaneously, each question consisting of a search and its subsearches.

## 1-8. OUTPUT BOUNDS

As shown in steps 2 and 4 of figure 1, this retrieval method also can accommodate bounds for the output of any one question, so that, for example, the inquirer may state that if there are less than ten references to the search subject (the class represented by the question), an additional set of keys should be searched as well; thus, if the file contains little on cats and dogs, what is there on pets? Or, the inquirer may state that if there are more than some arbitrary upper limit, say 100 references to the search subject, the machine should abandon the search and inform the human inquirer that he asked too comprehensive a question to be practical.

The method provides for the option that if subsearches and sub-subsearches are performed, a reference will be recorded for only the most specific level of pertinence (figure 1, step 2). For example, assume that a question calls for the descriptors

astronauts
gravity
moon
spaceships (A)

and the inquirer stipulates that a more general interest of his could be expressed by the descriptors

```
astronauts
gravity                                  (B)
spaceships
```

If a document contains all four of the descriptors called for in A, that document number is recorded as being pertinent for A, but not for B. Several reasons for adopting this option are discussed more fully later.

## 1-9. OPTIONAL OUTPUT INFORMATION

The output of this retrieval system can provide not only the material needed to satisfy the inquirer, but a record of the interrogation details which allow the system administrators to analyze its efficiency. For example, a record may be kept of what question was asked (what keys were used); how many subsearches the question entailed; what keys were contained in the subsearches; how many references were located for each search and subsearch; which of the references, after editing the search, were given to the inquirer; and so on.

By using this kind of information to analyze system function, one can determine such things as which input keys actually are used as output keys; in what combinations they are used; and with what degree of efficiency.

Such determinations are useful in revising retrieval-system components, or in revising human use of the components. Examples are (1) for making revisions to the input-output thesaurus, (2) for helping input-output staffs to achieve coincidence in their use of the thesaurus, (3) for helping the output staff to learn through experience how best to put questions to the computer system, and so on. The information also might be used for comparing one retrieval system with another to study their relative efficiencies.

## SECTION 2

## THE METHOD APPLIED AT ASTIA

### 2-1. INTRODUCTION

The Armed Services Technical Information Agency (ASTIA), Arlington, Virginia, is the first organization to apply the method described. At the time of this writing, the application still is in development. There are some recent publications by the ASTIA staff describing their automation activities.[1,2]

Figure 2 contains a block diagram with facsimile tape formats for the method as it will be used at ASTIA; figure 3 contains a flow chart* of the ASTIA application. The latter flow chart is more detailed than that in figure 1, but reflects the limitations of a particular set of parameters and a particular configuration of equipment.

### 2-2. EQUIPMENT

The computer system used at ASTIA is a Remington Rand Univac® Solid State 90 Computer, with tape components. Output is by means of a Univac® High-Speed Printer and a card punch.

---

[1] Heald, J. Heston, "Project MARS," Special Libraries 51 (1960), pp. 115-121.

[2] Automation of ASTIA: A Preliminary Report, ASTIA Document No. 227,000. Available on request.

* This flow chart and the detailed logic of the method as applied at ASTIA were supplied by J. K. Henderson, Richard Mayhew, and Arnold Dubinsky, of the Univac® Sales Office, Washington, D.C.

## 2-3. INPUT FILE

At first, the input file was on punched cards; the punched cards, however, were recently replaced by magnetic tape. The input file contains approximately 250,000 document descriptions, with an average of 8 descriptors per document. The document descriptions are controlled by a thesaurus, used also by reference librarians for transcribing an inquiry into machine language.

## 2-4. OPTIONS

A number of options (described earlier) concerning subsearches, upper and lower bounds for the number of references contained in answers, and so forth, have been incorporated into the ASTIA system. A search tape will be made up each day for an average of 50 questions, each question entailing up to 9 subsearches. A sample question form for guiding search and subsearch formulation is shown in figure 4.

The option has been adopted that for any question, a pertinent document number is recorded only once: at the level of greatest specificity. Reasons for adopting this option are given under heading 2-13.

## 2-5. PROGRAM PROCEDURE

As shown on the block diagram (figure 2) the program consists of four runs and three sorts. One of the runs, however, has two phases.

## 2-6. RUN 1 (SEQUENCE)

Punched input cards, each containing one search or subsearch, are sequence checked as they are read to insure that they are in request-number order. (This is important in run 3.)

The number of descriptors used by each search is counted; this count along with the request number is written on the F file.

As each descriptor is counted, it is written, with the request number, on the search file (S).

Certain validity checks, not shown in figures 2 or 3, also are made on each card.

The last step in run 1 is to sort the S file to descriptor order.

## 2-7. RUN 2, PHASE I (PULL)

The S-file descriptors are matched against those in the document file (master file D), which is in descriptor order. When a descriptor is found in the master file it is written on the G file along with all of the document numbers associated with it.

## 2-8. RUN 2, PHASE II (MERGE)

All G-file document numbers are examined one at a time, and written with their descriptors once for each request with which they occur in the R file.

The record file (R) is sorted to document- and request-number order.

2-9. RUN 3 (PARTITION)

Successive records in the R file are examined until a change in document and request numbers is encountered. Successive records containing identical document and request numbers are called a partition; the number of records in a partition is noted.

The request number is used as an entry to the F file to count the descriptors used by that search. If this count is the same as the number of records in the partition, the document number in that partition satisfies the search.

All document numbers which satisfy a search, along with the request numbers, are written on the T file.

After the partitioning is completed, the T file is sorted to request-number order.

2-10. RUN 4 (ANALYZE)

Run 4, the last, consists of the final analysis. If there are less than the optional number of references for each run (e) to level A for each question, then level-B hits are printed; otherwise they are discarded. In addition, a document number is recorded at the level of greatest specificity.

2-11. OUTPUT

At the end of the machine manipulations, the reference librarian receives an output of three cards. Cards 1 and 2 are 5- by 8-inch cards, card 3 a standard machine-punched card.

2-12. CARD 1

Card 1, shown in figure 5, is prepared by the high-speed printer; it identifies the search and gives the document numbers pertinent to that search or subsearch.

2-13. CARD 2

As was mentioned earlier, the option was adopted for the ASTIA application that for any question, a pertinent document number is recorded only at the level of greatest specificity. This option was chosen because card 2, shown in figure 6. is prepared for each pertinent reference; it identifies the document and search or subsearch, and gives the abstract of the document. If this option had not been adopted, many additional, time-consuming abstract copies would have been needed. Since only one copy of an abstract will be used for answering an inquiry under this option, subsearches may be used without checking the document numbers or abstracts for duplication.

2-14. CARD 3

The pertinent facts about a search are recorded in machine code on card 3, a punched card. These cards facilitate automatic compilation of reports about retrieval-system use and efficiency, as suggested earlier in this paper.

## CONCLUSIONS

The options provided in this generalized information-retrieval method (shown in figure 1) can be used to define the characteristics of a particular system so that it can be compared with another system defined according to the same terms.

Using the plan outlined in figures 1 and 4, system operators can decide whether the system is to answer 2 or 200 questions per day, whether they must be answered in minutes or hours, and whether the output will consist of reference numbers, original material, or photostats, and so forth. When these decisions have been made, the necessary staff and machine capacities can be determined.

The ASTIA-application flow chart (figure 3) may be helpful in automating future retrieval systems, even though different equipment or options are chosen.

Any input-output equipment can be used with the method—including equipment of the future.

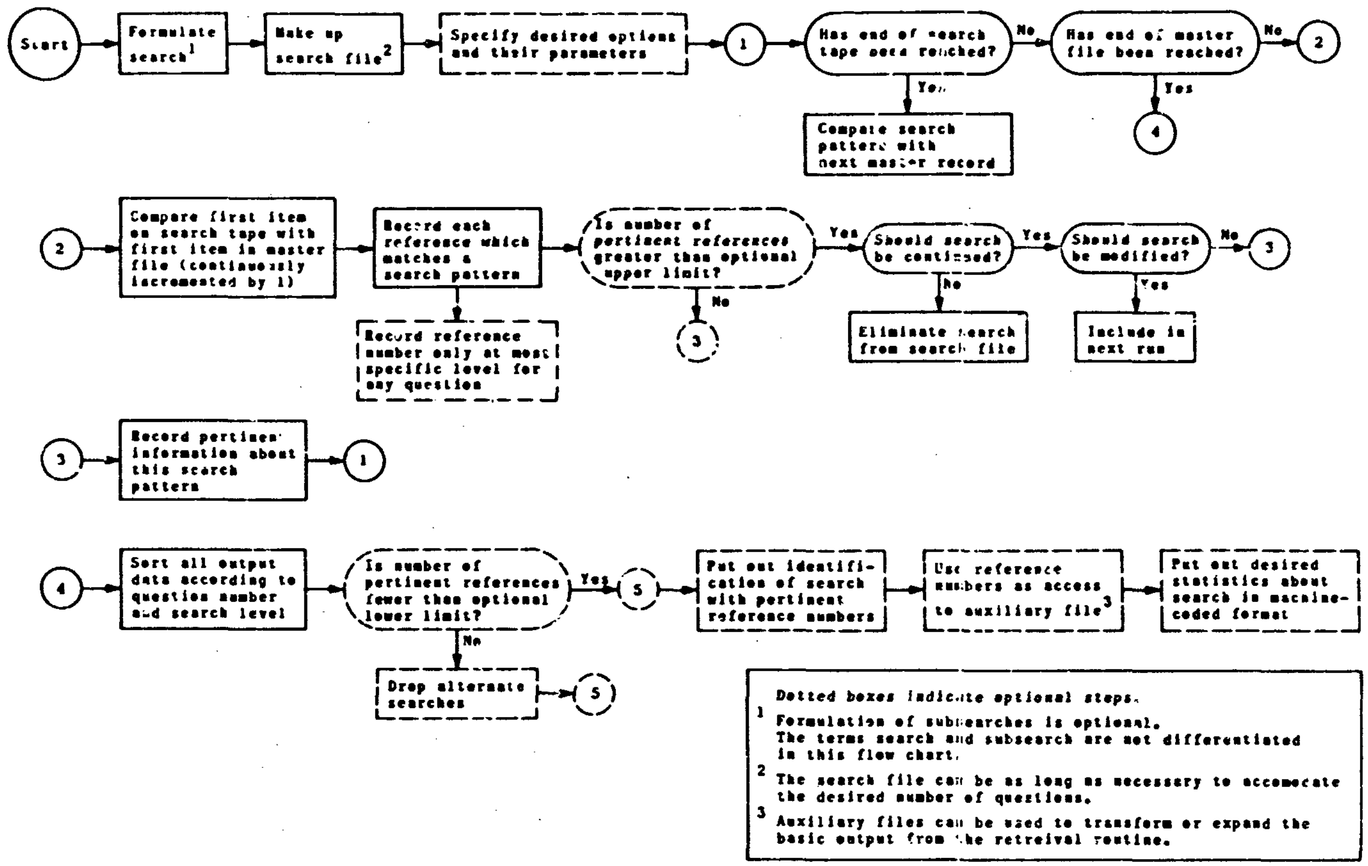
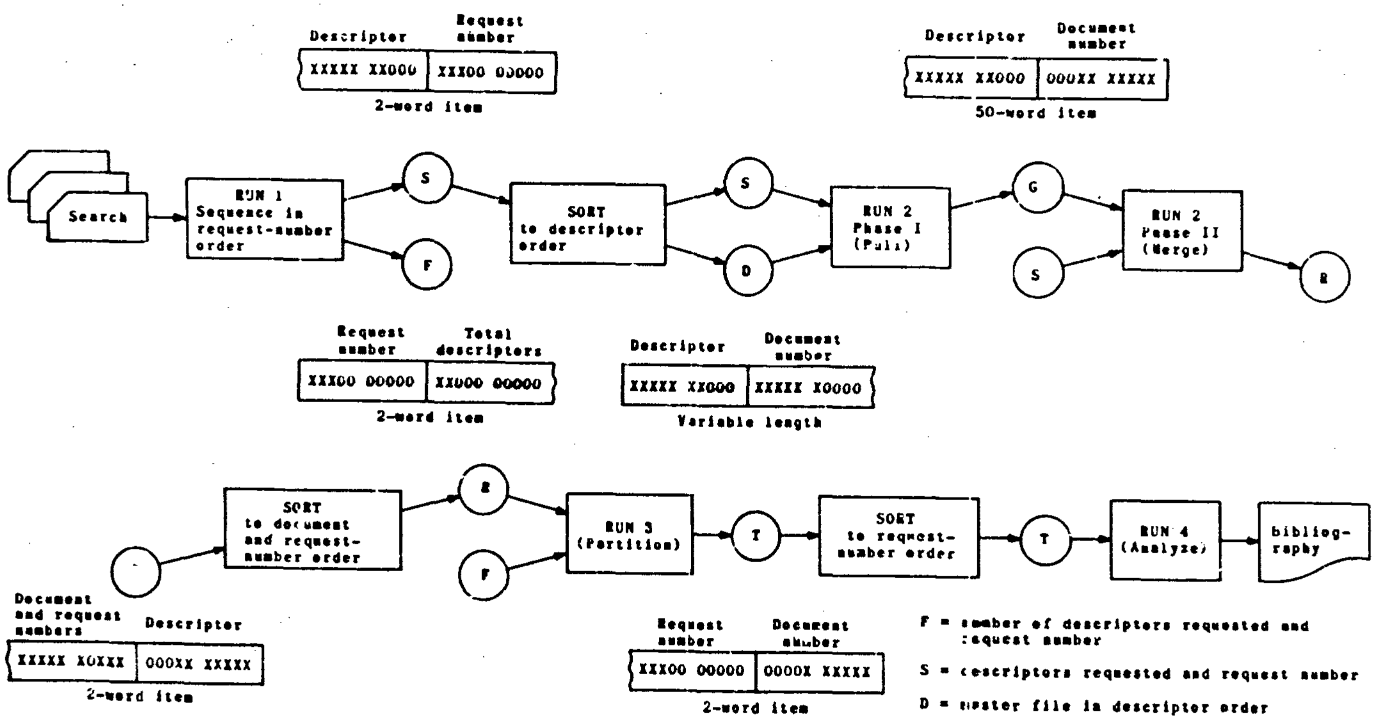



Figure 1. Generalized Flow Chart of Computer Program for Information Retrieval




F = number of descriptors requested and request number

S = descriptors requested and request number

D = master file in descriptor order

G = descriptors from master file matching those from S file and document numbers from master file

R = document numbers from G file and their descriptors (written once for each request with which they occur)

T = document numbers which satisfy search and request numbers

Figure 2. Block Diagram and Sample Tapes of Information-Retrieval Process at ASTIA

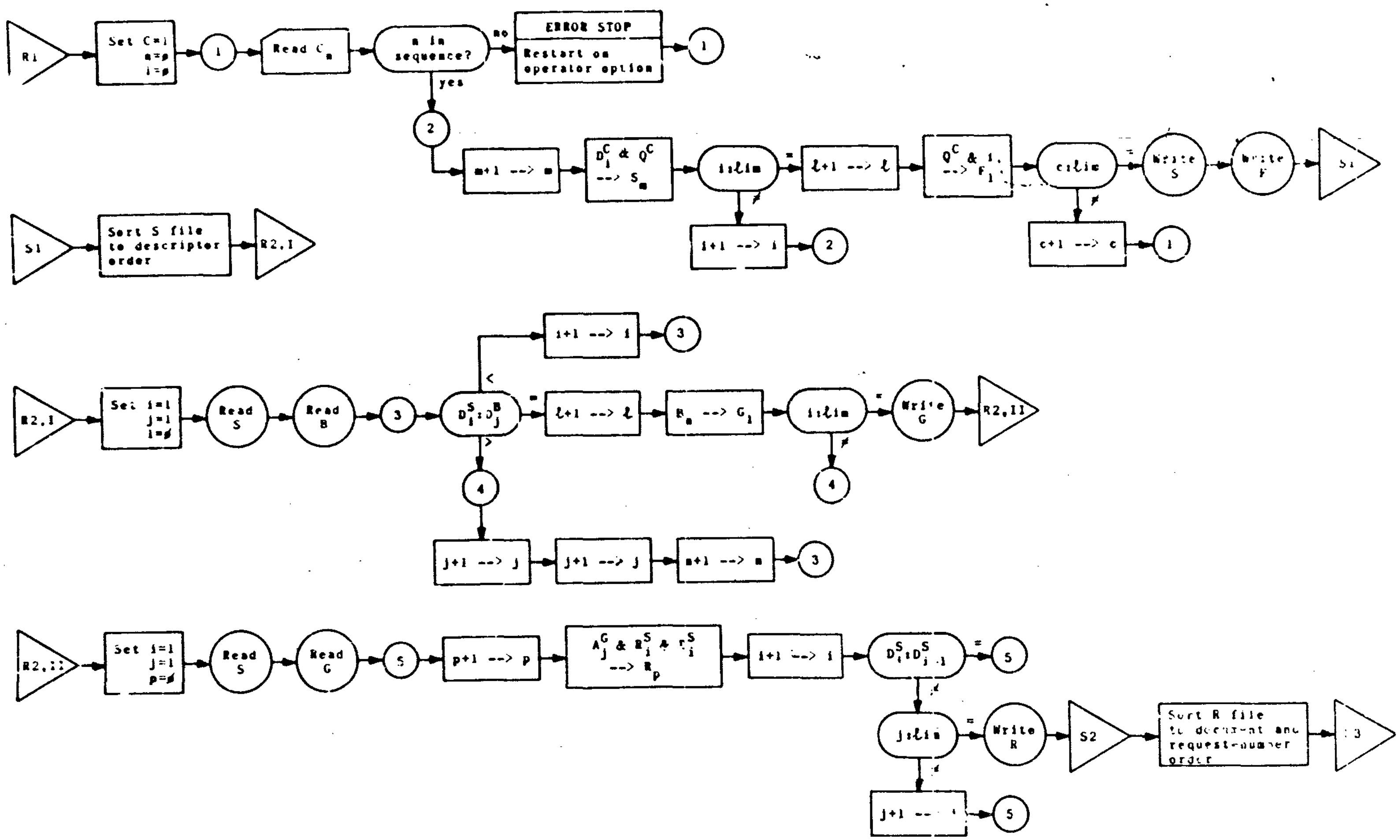
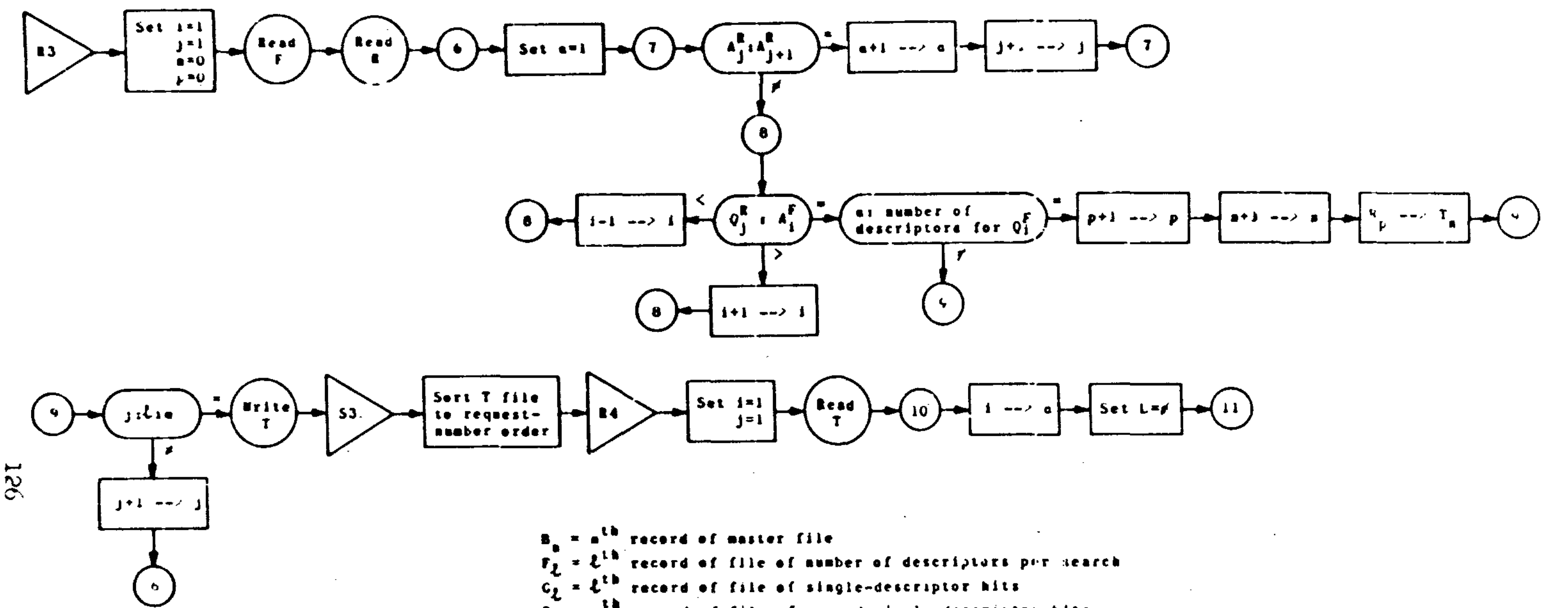



Figure 3. Flow Chart of Information-Retrieval Process at ASTIA




$B_n = n^{th}$ record of master file
$F_\ell = \ell^{th}$ record of file of number of descriptors per search
$G_\ell = \ell^{th}$ record of file of single-descriptor hits
$R_p = p^{th}$ record of file of merged single-descriptor hits
$S_m = m^{th}$ record of search file
$T_n = n^{th}$ record of file of true hits
$A^y_x = x^{th}$ document number from y file
$D^y_x = x^{th}$ descriptor from y file
$Q^y_x = x^{th}$ request number from y file
$\ell$ = number of answers at level 0
$\ell m$ is a bounded variable used to represent, symbolically, the number of times some particular event will occur.

Figure 3. Flow Chart of Information-Retrieval Process [illegible] ASTIA (Cont.)

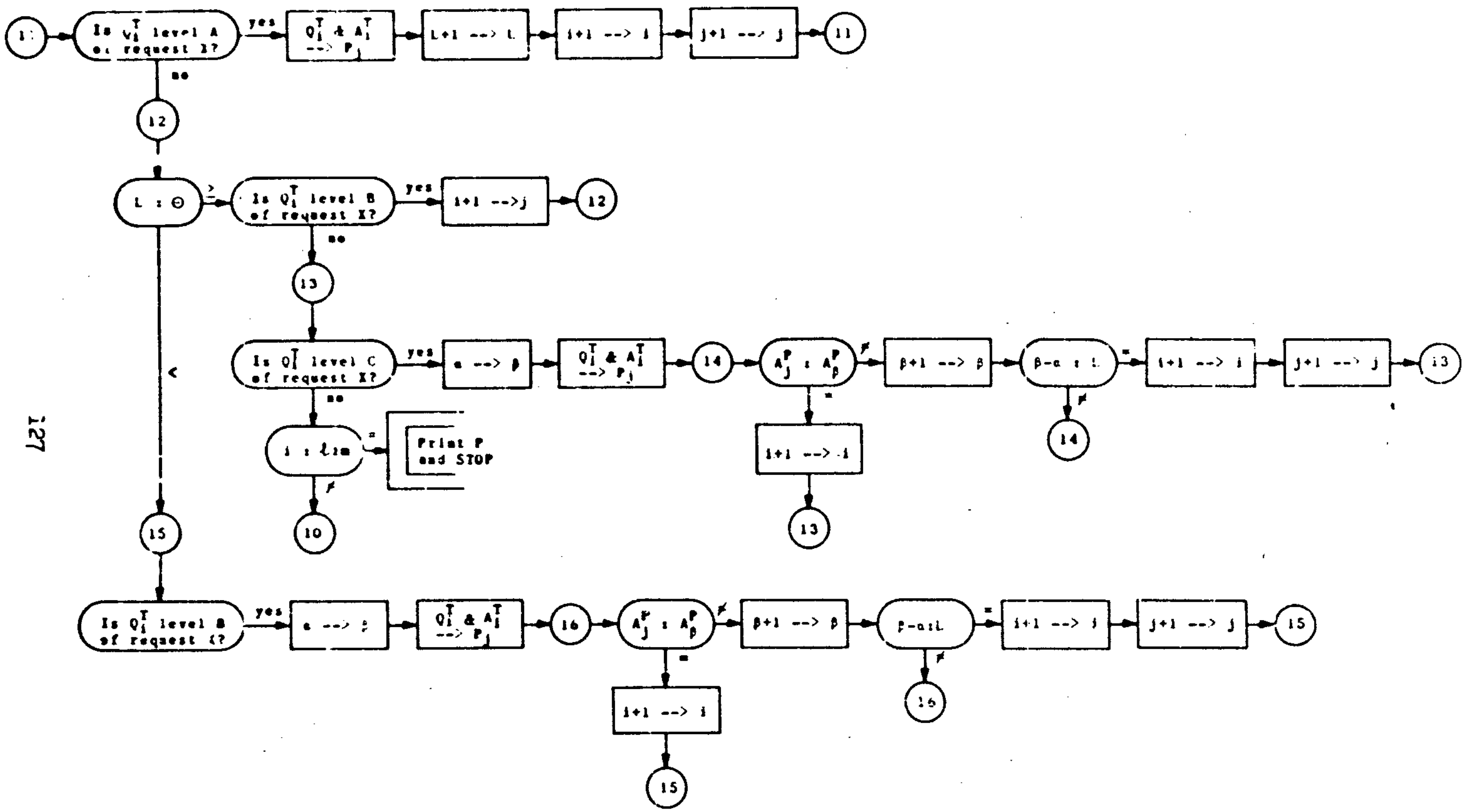



Figure 3. Flow Chart of Information-Retrieval Process at ASTIA (Cont.)

Question No. ____ Date ________________

A. Descriptors to be searched

Write descriptors here.

1. Astronautics
2. Moon
3. Space flight
4. Gravity
5.
6.

Order them here.

| | Descriptors | Codes |
|---|---|---|
| 100 | Astronautics | 610,000 |
| " | Gravity | 1,750,000 |
| " | Moon | 2,120,000 |
| " | Space flight | 6,241,000 |
| " | | |
| " | | |

B. If, and only if, there are less than 10 references to the above, which subsearches should be accepted? If the subsearches would be of interest even though there were more than 10 references pertaining to (A), list them under (C) instead of here.

| | Codes |
|---|---|
| 101 | 1,750,000 |
| " | 2,120,000 |
| " | |
| " | |
| " | |

| | Codes |
|---|---|
| 102 | 1,750,000 |
| " | 2,120,000 |
| " | |
| " | |
| " | |

| | Codes |
|---|---|
| __3 | |
| " | |
| " | |
| " | |
| " | |

| | Codes |
|---|---|
| __4 | |
| " | |
| " | |
| " | |
| " | |

List codes in ascending sequence.

C. Would any other combinations of the descriptors given in (A) be of interest? If so, give the code combinations desired. (limit of 5)

| | Codes |
|---|---|
| 105 | 610,000 |
| " | 1,750,000 |
| " | 6,241,000 |
| " | |
| " | |

| | Codes |
|---|---|
| __6 | |
| " | |
| " | |
| " | |
| " | |

| | Codes |
|---|---|
| __7 | |
| " | |
| " | |
| " | |
| " | |

| | Codes |
|---|---|
| __8 | |
| " | |
| " | |
| " | |
| " | |

| | Codes |
|---|---|
| __9 | |
| " | |
| " | |
| " | |
| " | |

Figure 4. Samy's Question Form for Guiding Search and Subsearch Formulation

```
(DATE) 11-17-61   (QUESTION NO.) 40   (LEVEL) U   (DESCRIPTORS) GERMFREE ANIMALS; DIET; GROWTH
221 330     246 118     450 315
227 010     246 914
229 414     246 917
229 417     247 008
229 418     247 008
235 018     249 316
235 019     249 707
240 117     249 718
```

Figure 5. Card 1, Output Showing Search Number and Pertinent-Document Numbers

AD-221 330 DIV. 16 (5 ANAL.: 80) LOBUND INST., U. NOTRE DAME, IND.

REARING GERMFREE CESAREAN-BORN RATS, MICE, AND RABBITS THROUGH WEANING. BY JULIAN R. PLEASANTS. (1959) (11) P INCL. ILLUS. TABLES. 16 REFS. (CONTRACT NONR-1623(04)) UNCLASSIFIED REPORT. REPRINT FROM ANNALS OF THE NEW YORK ACADEMY OF SCIENCES 78:116-126, 8 MAY 1959. DESCRIPTORS: GERM FREE ANIMALS; REARING; DIET; FEEDING; GROWTH PHYSIOLOGY; DIGESTIVE SYSTEM.

GERMFREE RATS, MICE, AND RABBITS WERE SUCCESSFULLY WEANED FROM HAND REARING BY RELATIVELY SIMPLE PROCEDURES. THE RATE OF SUCCESS WAS HIGH FOR RATS, MODERATE FOR RABBITS, AND LOW FOR SWISS MICE. THESE DIFFERENCES APPEARED TO REFLECT DIFFERING MARGINS OF SAFETY BETWEEN THE AMOUNTS OF DIET REQUIRED FOR GROWTH TO WEANING, AND THE AMOUNTS THAT COULD BE FED WITH LITTLE RISK OF INJURY TO THE DIGESTIVE OR RESPIRATORY TRACTS. ALTHOUGH GROWTH AND DEVELOPMENT WERE SUBNORMAL DURING THE SUCKLING PERIOD, THE DIET APPEARED TO BE SUFFICIENTLY BALANCED IN ITS INSUFFICIENCIES TO CREATE FEW CASES OF PERMANENT HANDICAP. WHILE MUCH CAN STILL BE DONE IN AN EMPIRICAL WAY TO IMPROVE INFANT GROWTH AND DEVELOPMENT, OPTIMAL RESULTS MAY DEMAND A THOROUGH UNDERSTANDING OF THE PHYSIOLOGY OF DIGESTION IN THE SUCKLING ANIMAL. (AUTHOR) (SEE ALSO AD-221 342)

Figure 6. Card 2, Output Showing Pertinent-Document Abstract

UNCLASSIFIED

UNCLASSIFIED